High Rate Signal Processing Schemes for Correlated Channels in 5G Networks

by

Sunil Dhakal

A dissertation submitted for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering.

University of Canterbury
Christchurch, New Zealand
February 2019

Abstract

Fifth generation (5G) cellular networks must support dense user environments where close proximity users and their associated highly correlated channel vectors are prevalent. Therefore, this thesis considers a multi-user multiple-input multiple-output (MU-MIMO) communication system, where both highly correlated users and semi-orthogonal users are present. Specifically, this thesis provides a performance analysis of several established and novel linear signal processing schemes for users with highly correlated channels. The main focus of this thesis is divided into two scenarios.

First, we have proposed two novel schemes namely decorrelating zero forcing (DZF) and hybrid zero forcing (HZF), which are robust in user environments where highly correlated users are present. DZF simply decorrelates the channels of two highly correlated users exploiting the advantage of mutual orthogonality of eigenvectors of highly correlated users, i.e., DZF employs the first and second eigenvectors while designing the precoders for two highly correlated users. DZF is as simple as conventional zero forcing (CZF), but achieves higher rates in highly correlated channel environments. Analysis and numerical results are presented for MU-MIMO systems demonstrating the impact of the DZF scheme, which improves user rates and provides fairness while scheduling the highly correlated users. Then, a more robust HZF scheme is designed by integrating the CZF and DZF schemes. In CZF, semi-orthogonal users are scheduled to achieve higher sum-rate, meaning that fairness among correlated users is compromised. HZF allows us to harness the advantages of both CZF and DZF, hence providing robustness against the joint scheduling of semi-orthogonal and highly correlated users with very little additional complexity.

Second, we studied the existing downlink non-orthogonal multiple access zero-

forcing beamforming (NOMA-ZFBF) system that exploits the high correlation between users. Furthermore, we propose a new power allocation scheme that ensures the weak user's quality of service (W-QoS) in a pair of strong and weak highly correlated users. Assuming an arbitrary correlation among the clustered users we derive the signal to noise ratio (SNR) and signal to interference plus noise ratio (SINR) expressions of clustered strong and weak users. These expressions contribute towards two important insights; a) the required correlation of clustered users and b) the required power differential between the highly correlated users in a cluster. Further, assuming perfectly correlated users we derive the exact SNR/SINR distribution of strong users, weak users and singletons. On the basis of these distributions we compute a closed-form approximation for the expected sum-rate of the NOMA system that gives an upper-bound on system performance. This is useful to understand how the correlation factor affects the clustering rate and how close partially correlated channels can get to the perfectly correlated bound. Numerical results confirm the reliability of the proposed W-QoS NOMA based techniques and lead to the important insights required to understand the performance of NOMA in 5G cellular communication.

Finally, the variance of the interference power between two user channels in a downlink Rician fading channel with spatial correlation at the Base Station (BS) is analyzed. We show that high variance is beneficial to NOMA so that NOMA performance is directly linked to the variance of the interference power. This leads to three key properties linking channel statistics to NOMA potential, namely; a) similar channel statistics increase the variance, b) equal correlation matrices increase the variance and c) similar line of sight (LoS) directions increase the variance. Comparison results presented between correlated and uncorrelated Rayleigh and Rician channels suggest that highest variance can be obtained in correlated Rayleigh channels with a narrow angular spread which yields higher cluster percentage, inter-cluster correlation, and cluster-rate.

In summary, we can conclude that the DZF scheme can achieve a higher rate than CZF in highly correlated user environments and HZF is robust to the user environment where both similar channel users and orthogonal channel users co-exist. Furthermore,

when ZFBF is employed in NOMA systems, then a power differential among the highly correlated clustered users is not required. Finally, correlated Rayleigh channels having similar channel statistics are most suitable for NOMA-ZFBF due to the higher variance it provides for the interference power.

Acknowledgements

Firstly, I would like to express my gratitude to my supervisors Prof. Philippa Martin and Prof. Peter J. Smith, for their continuous support and guidance throughout my Ph.D. journey. I am grateful to Prof. Philippa Martin for giving me an opportunity to pursue the Ph.D. degree at the University of Canterbury. I have highly benefited from creative research ideas from Prof. Peter J. Smith and constructive feedbacks from both of my supervisors. The contributions in this thesis would not have been the same without them.

A special thank goes to my friends Diwakar Bhujel, Subodh Pudasaini, Subodh Dhakal, Salina Dhakal, Jalesh Devkota, Sheetal Sharma, Narayan Nepal, Dipendra Ghimire, and Deepak Chouhan for being there and helping me in every ups and downs in these four years of my Ph.D. journey. Also, I would like to take this opportunity to thank all my friends from the University of Canterbury and Christchurch.

I would like to acknowledge with gratitude, the love and support of my family - my parents, Harisaran Dhakal and Lalita Devi Dhakal; my brother, Suman and my sister, Sonu. I would also like to thank my extended family- sisters-in-law, Sajita Dawadi and Aarchana Karki; nephew, Asutosh; father-in-law, Jagganath Karki; mother-in-law, Rati Karki and brothers-in-law, Jagdish Karki and Santosh Pudasaini. They all have been supportive in every way possible.

Most importantly, I owe my deepest gratitude towards my better half, Rajani Karki, for her eternal support and for being so understanding in these years. This thesis would not have been possible without her patience and sacrifice. I also dedicate this thesis to my two lovely daughters, Aiesha and Azalea, who are the bundles of joy of my life.

Contents

List of Figures

List of Tables

Abbreviations

3G	Third generation.
3GPP	3rd generation partnership project
4G	Fourth generation.
5G	Fifth generation.
AOA	Angle-of-arrival.
AOD	Angle-of-departure.
AWGN	Additive white Gaussian noise.
BF	Beamforming.
BS	Base station.
CDF	Cumulative distribution function.
CSI	Channel state information.
CSIT	Channel state information at transmitter.
CUC	Correlated user clustering.
CZF	Conventional zero forcing.
DZF	Decorrelating zero forcing.

HZF	Hybrid zero forcing.
I.I.D.	Independent and identically distributed.
ICI	Inter-cluster interference.
IUI	Inter-user interference.
LOS	Line of sight.
LTE	Long term evolution.
MIMO	Multiple-input multiple-output.
MMSE	Minimum mean squared error.
mmWave	Millimeter-wave.
MU-MIMO	Multiuser multiple-input multiple-output.
NLOS	Non-line-of-sight.
NOMA	Non-orthogonal multiple access.
OMA	Orthogonal multiple access.
QoS	Quality of service.
SC	Superposition coding.
SIC	Successive interference cancellation.
SISO	Single-input single-output.
SINR	Signal-to-interference-plus-noise-ratio.
SNR	Signal-to-noise-ratio.
SUS	Semiorthogonal user selection.

SVD	Singular value decomposition.
TDMA	Time division multiple access.
UE	User equipment.
ULA	Uniform linear array.
UM	User matching.
ZF	Zero forcing.
ZFBF	Zero forcing beamforming.

Notation

$\mathbf{I}_M$	$M \times M$ identity matrix.
$(\cdot)^T$	Matrix or vector transpose.
$(\cdot)^H$	Conjugate transpose.
$(\cdot)^{-1}$	Inverse.
$(\cdot)^\dagger$	Pseudo inverse.
$\lvert \cdot \rvert$	Absolute value.
$\lvert\lvert \cdot \rvert\rvert$	Norm of a matrix or vector.
$E(\cdot)$	Expectation.
$\mathrm{Var}(\cdot)$	Variance.

Chapter 1

Introduction

In modern history one of the most successful technological innovations is cellular communication. Today, it has become a part of life for 5 billion people [1]. Continuous research and development during the last decade has led to a paradigm shift from voice-centric 2nd and 3rd generation cellular communication towards data-centric 4th generation (4G) cellular communication [2]. Current 4G cellular communication systems, such as long term evolution (LTE) [3], LTE-Advanced (LTE-A) [4] and worldwide interoperability for microwave access (WiMax) [5], are providing seamless connectivity over the air-interface. In 4G, cellular providers are employing advanced radio technologies, such as orthogonal frequency division multiplexing (OFDM), code division multiple access (CDMA), multiple-input multiple-output (MIMO), multi-user diversity, to fulfill the capacity demand, connectivity and better quality of service (QoS) requirements [6].

On the other hand, given the explosive growth of mobile phone users over the years along with the steadily increasing use of video streaming in smart-phones and tablets, it has been predicted that the 5th generation (5G) cellular communication systems need to fulfill a 1000 fold increase in capacity demand along with high data rate, low latency and ubiquitous connectivity [7] [8]. To fulfill these demands, 5G should differ fundamentally from its predecessors encouraging a series of groundbreaking innovations [2]. Out of many research areas in 5G, some researchers are working on the efficient use of spectrum using MIMO system [9], advanced coding and modulation

schemes [10], while some are researching new frequency bands, such as millimeter wave (mmWave) [11] [12] [13], that can boost the wireless connectivity and data rate. To demonstrate the QoS and service availability, prototypes are being tested in outdoor environments for both Line of sight (LoS) and Non-line of sight (NLoS) sites [2] [14] [15] and propagation pathloss models are being designed [16] [17] [18]. Moreover, MIMO and beamforming (BF) solutions for both microwave and mmWave systems are being investigated for 5G. Therefore, new technological solutions are evolving to ensure the sustainability of mobile communication services over the next generations [6].

1.1 Motivation

The pioneering work of Winters [19], Foschini [20], Telatar [21], Foschini and Gans [22] have predicted that MIMO systems can provide remarkable spectral efficiencies for wireless systems [23]. Extensive research on MIMO systems over the past decade have resulted in its deployment in current wireless systems [24] [25]. Further, it has been recognized as a key enabling technology for 5G cellular networks due to its potential to meet the demands of higher data rates, improved spectral efficiency, and better QoS [26] [27].

In general, MIMO systems exploit the spatial domain adding antennas at transmitter and/or receiver to achieve spatial diversity gain, spatial multiplexing gain and interference reduction [28] [29]. In spatial diversity, multiple copies of the same signal are transmitted from spatially separated antennas. This increases the reliability of reception, by increasing the probability that at least one copy of the signal out of the multiple copies received is not experiencing a deep fade [28]. Furthermore, in spatial multiplexing, independent data streams are transmitted from each antenna and this increases the capacity of the wireless network. These data streams are then separated at the receiver using appropriate signal processing schemes such as minimum mean squared error (MMSE) and zero-forcing (ZF) [30] [31] [32] [33]. Despite its sub-optimal performance, ZF has gained popularity due to its lower complexity

with acceptable performance compared to other schemes. The main advantage of ZF is that it completely eliminates the inter-user interference [34] [35] with perfect channel state information. However, the performance of ZF degrades when the users' channels are correlated [36] [37] [38] due to noise inflation during channel inversion. As a result, most approaches to ZF in multiuser-MIMO systems target the selection of users with strong, uncorrelated channels [33] [39].

On the other hand, emerging papers in 5G are discussing dense user environments and ultra-high connectivity where close-proximity users and their associated highly correlated channel vectors are becoming more prevalent [40] [27] [41] [42]. Moreover, industrial field trials have found that typical linear processing schemes are poor at separating closely spaced users [43] [44]. Instead, they employed more complex non-linear techniques, Tomlison-Harashima precoding (THP) and a hybrid precoding which combine eigen ZF (EZF) and THP [44]. Hence, in these scenarios, it is reasonable to implement techniques such as spatial separation of users via massive base station (BS) arrays [41] or non-orthogonal multiple access-beamforming (NOMA-BF) [40] [27] [45] [46] [47]. In massive MIMO technology, large numbers of antennas are deployed that rely on phase-coherent techniques but employ very simple processing of signals at the BS [9]. Asymptotic orthogonality of channel vectors associated with different users are the key to the performance of massive MIMO systems [48] [49]. Despite the potential of massive MIMO systems, it also comes with several issues such as pilot contamination, computational complexity, channel estimation overheads and energy consumption [50] [51] [52]. These issues are being investigated resulting in potential solutions. However, in this thesis rather than looking at massive MIMO systems, we concentrate on NOMA zero-forcing beam-forming (NOMA-ZFBF) techniques. In a conventional multi-user ZFBF system BF vectors are designed such that the beams are orthogonal to the other supported users' beams. This results in a complete elimination of the interference from other users thereby increasing the sum capacity of the system. Further, a single BF vector supports a user at a time. However, NOMA-BF is designed such that a single BF vector supports multiple users. Thus, the number of supported users in one time slot increases which in turn enhances

the sum-capacity of the system [40].

Over the last few years, NOMA has attracted a lot of attention as a key enabling technology for 5G networks. NOMA is a power domain multiplexing technique where multiple users can be multiplexed at different power levels using the same time/frequency/code [27] [53]. In [40], multi-user ZFBF for downlink MIMO-NOMA systems was proposed where the users are grouped into a cluster based on their channel correlation and gain differences. Then a single beam is employed to transmit signals to the highly correlated users in the cluster[1]. For the NOMA-ZFBF system in [40], it was shown that clustering highly correlated users is more likely to suppress the inter-cluster interference at the weak user and the inter-user interference in a cluster. A similar NOMA-ZFBF approach has been discussed in [26] [27], where users with distinct channel gains are paired in a cluster. Further, for the MIMO-NOMA system in [26] it was shown that the performance gain of NOMA increases when the users' effective channel gains in a cluster are more different, i.e., a bigger difference in the channel gains of the clustered users increases the performance gain of NOMA systems[2].

Therefore, understanding the performance of multiuser-MIMO systems for environments where users have similar channels or highly correlated channels is of growing importance. Furthermore, it is important to understand the performance gain of NOMA systems in different fading environments which allows researchers to explore different techniques in a more informed manner, especially in MU-MIMO systems. Analyzing different systems in different settings broadens our understanding of the system and innovative solutions result [54].

[1]In a cluster, users with strong channel conditions and weak channel conditions are referred as strong/near and weak/far users, respectively.

[2][26] has evaluated that the performance gain of NOMA increases when the users' effective channel gains in a cluster are distinct. However, with NOMA-ZFBF in chapter 5, we show that the performance gain of NOMA can increase when the weak users' channel gain is closer to the strong user's gain.

1.2 Thesis Outline and Contributions

In this thesis, we will present novel receiver processing schemes for environments with highly correlated user channels. Moreover, we will study and analyze the MU MIMO-NOMA systems in different fading channels.

Chapter 2

Chapter 2 provides the relevant concepts and techniques used throughout the thesis. Fundamentals of wireless channels, fading channel models, MU-MIMO systems, NOMA systems and precoding techniques are described to understand the basic physical layer concept of wireless cellular communication. Moreover, correlation among the users is discussed, which is the foundation and main motivation of this thesis.

Chapter 3

In chapter 3, we propose a novel transmit/receive processing scheme denoted decorrelating zero-forcing (DZF). The scheme improves downlink ZF performance over MU-MIMO channels when highly correlated users are present. In particular, DZF groups correlated users into pairs as in a NOMA system, where the stronger user is labelled "near" and the weaker user, "far". For the far users and users not paired, the proposed design is similar to conventional-ZF (CZF) in that the base station performs traditional ZF and the users employ the leading left singular vectors of the channel for processing. In contrast, the near users employ the second strongest left singular vector which decorrelates their effective channel with the paired far user. This reduces the noise inflation inherent in ZF for correlated channels. From simulations and analysis we demonstrate that DZF has both rate and fairness advantages over CZF when highly correlated users are present.

Chapter 4

This chapter extends the potential of the DZF scheme proposed in chapter 3 integrating DZF with the CZF scheme. DZF can schedule highly correlated users without

degrading the system performance while CZF schemes schedule semi-orthogonal users to yield a performance gain. Therefore, combining the DZF and CZF schemes can provide a robust scheme, referred to as hybrid zero-forcing (HZF), that can leverage the potential of both schemes. HZF is a simple yet an efficient algorithm that provides fairness while scheduling the users and achieves performance gain.

Chapter 5

In chapter 4, we discuss a HZF scheme that is beneficial in the scenarios where both semi-orthogonal and highly correlated users are present. Considering a similar scenario, chapter 5 studies the NOMA-ZFBF technique, where the same beamforming vector is employed to highly correlated users and ZFBF is used to nullify the interference between the separate beams.

Specifically, chapter 5 provides an efficient algorithm that ensures the weak users' QoS in a cluster of highly correlated users in NOMA-ZFBF systems. Furthermore, we derive and analyze the exact signal to noise ratio (SNR) or signal to interference plus noise ratio (SINR) distributions of singletons [3] and highly correlated strong and weak users in a cluster. The key contribution of this investigation is that our approach doesn't require a power differential between the clustered users. In fact, both channels being strong helps clustering and overall cluster rate. Moreover, we derive a novel closed form upper bound sum-rate for the NOMA system.

Note that there are clear links between HZF and NOMA schemes although the focus is quite different. With NOMA the focus is on increased capacity delivered by using the same BF vector to highly correlated users and using successive interference cancellation methods and power allocation to separate users. With HZF, the transceivers are much simpler and instead of deliberately scheduling correlated users (as in NOMA), the scheme allows ZF to operate successfully in all scenarios, even in the presence of highly correlated users.

[3] In a NOMA system, singletons are the users that cannot be grouped into a highly correlated pair.

Chapter 6

The NOMA-ZFBF systems discussed in chapter 5 require both semi-orthogonal users (SU) in beams and highly correlated (HC) users within a beam to maximize the spectral efficiency of the system. Hence, channel models where different users are likely to experience both similar and near-orthogonal channel vectors are likely to give good NOMA performance. In order to explore this, we analyze the variance of the interference power between two users, where a large variance indicates this channel behavior is more likely. Moreover, we have compared the Rician and Rayleigh channels in correlated and independent and identically distributed (I.I.D.) environments in order to provide some new insights into the role of LoS and correlated components with respect to NOMA beam formation and clustering of HC users. Analysis and numerical results show that correlated channels with similar channel statistics result in a higher variance that in turn increases the cluster percentage and enhances the cluster rate.

Chapter 7

Finally in this chapter we conclude the thesis summarizing the key contributions. Specifically, this chapter includes future research directions and discusses the potential of our research in mmWave communication.

1.3 List of Publications

1. **S. Dhakal**, P.A. Martin, P. J. Smith, "Decorrelating Zero-Forcing for Correlated MU-MIMO Channels". (To be submitted).

2. **S. Dhakal**, P.A. Martin, P. J. Smith, "Hybrid Zero-Forcing for Correlated and Semi-orthogonal MU-MIMO Channels". (To be submitted).

3. **S. Dhakal**, P.A. Martin, P. J. Smith, "NOMA with Guaranteed Weak User QoS: Design and Analysis", (Accepted in IEEE Access).

4. **S. Dhakal**, P.A. Martin, P. J. Smith, "On The Impact of Line-of-Sight and Spatial Correlation on NOMA Performance", in *Proc. IEEE Vehicular Technology Conference (VTC 2018 Fall)*, Chicago, USA, Aug. 2018.

Publications 1 and 2 are related to Chaps. 3 and 4 of this thesis. Publication 3 is related to Chap. 5 and publication 4 is related to Chap. 6.

Chapter 2

Background

This thesis relies on some existing wireless standards, parameters and mathematical tools in order to analyze the physical layer of cellular communications. Hence, this chapter will provide a review of the concepts and techniques used throughout the thesis. First, we briefly describe the channel model, MU-MIMO system model and precoding techniques. Then orthogonal multiple access (OMA) and non-orthogonal multiples access techniques (NOMA) are described. Finally, the correlation between the users and their effects using MIMO-ZFBF are discussed.

2.1 Channel model

In wireless communication, radio signals are transmitted from a transmitter over a wireless channel to a receiver. Such channels encounter multiple radio-interfering objects, such as trees, poles, buildings and vehicles. Hence, the channel is impaired due to absorption, reflection, diffraction and obstruction from these objects which is commonly known as shadowing [23]. Additionally, the channel power decays over distance which is known as path loss. The received power variation due to path loss and shadowing is jointly known as a large scale propagation effect [23] or large scale fading [34] [52]. Moreover, when the signals are reflected or diffracted from the physical obstructions multiple copies of the transmitted signal arrive at the receiver via different channel paths. This results in constructive or destructive addition of

the multiple channel components which is known as small scale propagation effects or small scale fading [23] [55] [50] [56]. The channel model considering both small scale and large scale propagation effects can be expressed as[1] [52] [50]

$$h = \sqrt{\beta} g, \tag{2.1}$$

where β is due to large scale fading, g is small scale fading and h represents the overall fading of a single-input single-output (SISO) wireless channel.

2.1.1 Large scale fading

As the large scale fading is due to the pathloss and shadowing effects, it is difficult to obtain a single model that can accurately characterize both pathloss and shadowing effects across a range of different environments. Different analytical models or empirical measurement based models are available for different propagation environments [57] [58]. In this thesis, we design β using a simple model defined by [59]

$$\beta = A\xi(r_0/r)^{-\gamma}, \tag{2.2}$$

where A is the unit-less constant for the geometric attenuation at a reference distance of r_0, assuming far field omni-directional antennas. r is the actual distance between the transmitter and receiver and γ is the attenuation exponent. Further, the effects of large obstacles are often statistically modeled as a log-normal random variable [57]

$$\xi = 10^{(X/10)}, \tag{2.3}$$

where X depends upon the shadow fading standard deviation (σ_{sf}), i.e., $X \sim \mathcal{N}(0, \sigma_{sf}^2)$.

2.1.2 Small scale fading

As discussed above, the multipath components of a channel result in constructive or destructive interference at the receiver. In addition, these multipath components

[1] Note that this can also be referred as time invariant frequency flat fading model.

differ with the environment and vary across time, frequency and spatial domains as detailed below [55].

- Time variation arises due to Doppler effect, i.e., either the transmitter or receiver is moving, hence the location of the reflectors are changing over time.
- Frequency variation is due to the several delayed and scaled versions of the signals arriving at the receiver.
- The spatial diversity can be realized using several antenna elements at receiver and/or transmitter side, which is further described in Sec. 2.1.3.

The variation of h depends on the channel models and parameters employed. Some of the small scale fading channel models used in this thesis are described below.

Rayleigh fading channel

A Rayleigh fading channel is the simplest channel model which is often used to model general propagation scenarios which are rich in scattering and do not contain a LoS component [30] [60]. A simple Rayleigh fading channel can be expressed as

$$g = \frac{1}{\sqrt{2}}(a + ib), \tag{2.4}$$

where a and b are independent Gaussian variables with zero mean and unit variance [52]. Hence, g can be written as

$$g \sim \mathcal{CN}(0, 1). \tag{2.5}$$

Furthermore, $E[|g|^2] = 1$ which gives $E[|h|^2] = \beta$, which is also known as the link gain or channel strength [23] [55].

Uncorrelated and correlated Rayleigh fading channel

To realize channel models in a multiple transmit-receive antenna environment and to discuss correlation among the transmit antennas at a BS, we consider a downlink

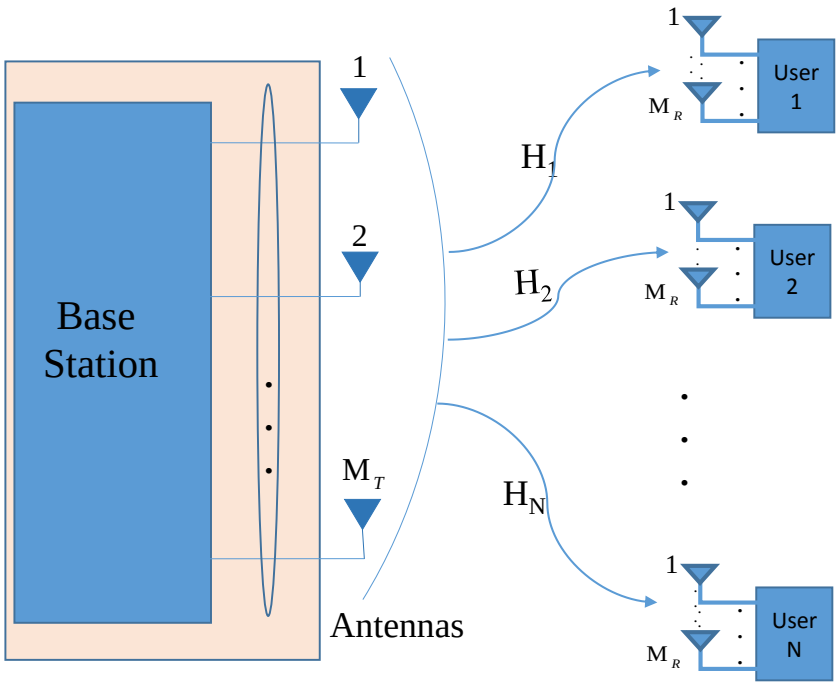


Figure 2-1: A MU-MIMO downlink system.

MU-MIMO system, where a total of M_T transmitting antennas at the BS serves the N user equipments (UEs), each with M_R receiving antennas which is shown in Fig. 2-1[2]. In the figure $\mathbf{H}_i$ is a $M_R \times M_T$ channel matrix for $i = 1, 2, \ldots, N$ which completely describes the fading characteristics of the entire MIMO link of the i^{th} user. The MIMO channel can then be described as

$$\mathbf{H}_i = \sqrt{\beta_i} \begin{bmatrix} g_{11} & \cdots & g_{1M_T} \\ \vdots & \ddots & \vdots \\ g_{M_R1} & \cdots & g_{M_RM_T} \end{bmatrix}, \tag{2.6}$$

where g_{xy} is an element of channel matrix $\mathbf{G}_i$ from the y^{th} BS transmit antenna to the x^{th} receive antenna of the i^{th} user. We assume that the entries of $\mathbf{G}_i$, satisfy $E[|g_{xy}|^2] = 1$, which can be written as $g_{xy} \sim \mathcal{CN}(0, 1)$ for a Rayleigh fading channel. Therefore, g_{xy} is called the normalized channel coefficient [54].

To distinguish the different channel models, the independent and identically dis-

[2]The part on MU-MIMO channels is included for completeness only.

tributed (I.I.D.) Rayleigh fading channel model is represented as $\hat{\mathbf{H}}_i$ [3]. Further, this model is used as a baseline throughout the thesis for evaluation of the considered cellular systems unless otherwise specified. If the transmit/receive antennas and the wireless channel do not introduce correlation, then the channel matrix is spatially white, which is known as an uncorrelated or I.I.D. Rayleigh channel [29]. Thus, the channel model in (2.1) can be rewritten as

$$\hat{\mathbf{H}}_i = \sqrt{\beta_i}\hat{\mathbf{G}}_i, \tag{2.7}$$

where $\hat{\mathbf{G}}_i$ denotes the $M_R \times M_T$ normalized downlink channel matrix from BS to the i^{th} UE, for $i = 1, 2, \ldots, N$.

The I.I.D. Rayleigh fading model is reasonable when there is no spatial correlation at both transmit and receiving sides of the link. However, the limited distance between the antenna elements or a small angular spread at either/both the TX and RX can cause spatially correlated channels. The Kronecker model[4] is often used while modelling spatial correlation at either/both the TX and RX [61]. Kronecker model is given as [28, p. 90]

$$\mathbf{R}_h = \mathbf{R}_t \otimes \mathbf{R}_r, \tag{2.8}$$

where $\mathbf{R}_h$ is the channel covariance of the $M_R M_T$ scalar channels between the M_T transmit and M_R receive antennas of $\mathbf{H}_i$ in (2.6). $\mathbf{R}_t$ is the covariance of the scalar channels seen from M_T transmit antennas to a single receive antenna (corresponding to row of $\mathbf{H}_i$) is the same for any receive antenna. Similarly, the covariance of the scalar channels seen from a single transmit antenna to all M_R receive antennas (corresponding to column of $\mathbf{H}_i$) is equal to $\mathbf{R}_r$ ($\mathbf{M}_R\mathbf{M}_R$). And $\otimes$ represents Kronecker product.

In this thesis we only consider spatial correlation at the BS side because the users may only have a few antennas and are often in a rich scattering environment. Thus,

[3] $\hat{\mathbf{H}}$ is mainly used in chapters 4 and 6, where we compare I.I.D. Rayleigh with other channels. In chapters 3 and 5, $\mathbf{H}$ also represents the I.I.D. Rayleigh channel.

[4] To understand the Kronecker model and the impact of channel correlation on the precoding scheme please refer to [28, p. 90, p. 110, p. 113].

the correlated Rayleigh fading channel with spatial correlation at the transmitter can be written as

$$\mathbf{H}_i = \hat{\mathbf{H}}_i \mathbf{R}_i^{1/2}, \tag{2.9}$$

where $\hat{\mathbf{H}}_i$ is defined in (2.7) and $\mathbf{R}_i$ (same as $\mathbf{R}_t$ in 2.8) is the $M_T \times M_T$ transmit spatial correlation matrix which is unique to each user. We will discuss the spatial correlation further in Sec. 2.1.3.

Uncorrelated and correlated Rician fading channel

A Rician fading channel considers the LoS component between the transmitter and the receiver. An important parameter in the Rician channel is the Rician factor (K), which is the ratio of the powers in the LoS to the diffuse components [62]. Further, considering both the LoS component and the scattering component the independent Rician channel can be modeled as

$$\mathbf{H}_i = \sqrt{\beta_i}\left(\sqrt{\frac{K}{K+1}}\bar{\mathbf{G}}_i + \sqrt{\frac{1}{K+1}}\hat{\mathbf{G}}_i\right), \tag{2.10}$$

where $\bar{\mathbf{G}}_i$, $\hat{\mathbf{G}}_i$ are the specular (LoS) and diffuse (scattered) components of the channel assuming uncorrelated fading, respectively. Assuming a uniform linear array (ULA), the LOS component is $\bar{\mathbf{G}}_i = [1, e^{j2\pi d \sin\bar{\phi}}, \cdots, e^{j2\pi(M-1)d\sin\bar{\phi}}]$, where $\bar{\phi}$ is the angle-of-departure (AoD) of the specular component and d is the inter-element spacing in wavelengths.

In uncorrelated Rayleigh fading we have $g_{xy} \sim \mathcal{CN}(0,1)$, where g_{xy} is an element of $\mathbf{G}_i$. However, in Rician fading,

$$g_{xy} \sim \mathcal{CN}\left(\sqrt{\frac{K}{K+1}}A, \sqrt{\frac{1}{K+1}}\right). \tag{2.11}$$

For the normalized Rician fading channel, $|A|^2 = 1$, hence $E[|h_{xy}|^2] = \beta_i$, where h_{xy} is an element of $\mathbf{H}_i$. Thus, the power of the LoS component is given by $\frac{\beta K}{K+1}$ [55].

Similar to the correlated Rayleigh fading channel, a spatially correlated Rician

fading channel is due to the limited distance between the antenna elements or a restricted angular spread at either/both the TX and RX. Therefore, the correlated Rician fading channel can be written as

$$\mathbf{H}_i = \sqrt{\beta_i}\left(\sqrt{\frac{K}{K+1}}\bar{\mathbf{G}}_i + \sqrt{\frac{1}{K+1}}\hat{\mathbf{G}}_i\mathbf{R}^{1/2}\right), \tag{2.12}$$

where $\mathbf{R}$ is the described in section 2.1.3.

2.1.3 Spatial fading correlation

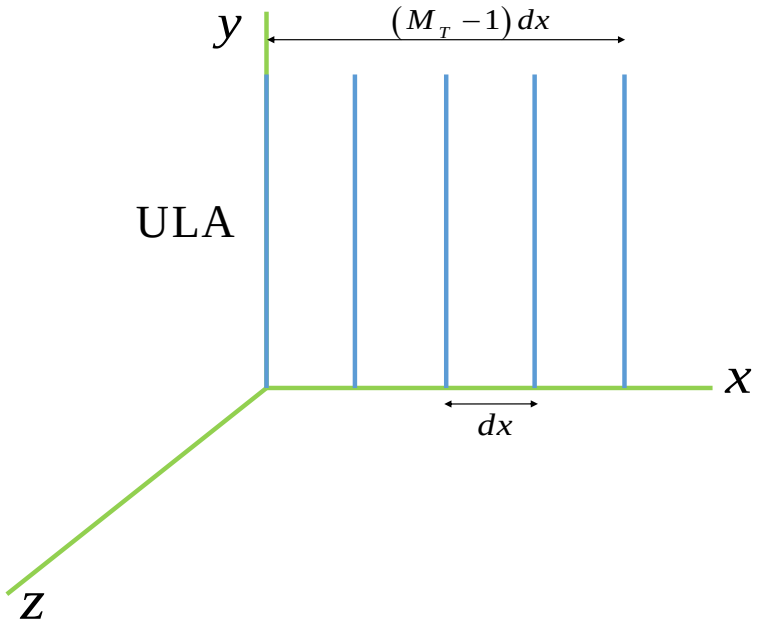


Figure 2-2: ULA antenna topology with M_T antennas on the x-axis with dx inter-antenna spacing.

In general, spatial correlation exists between the elements of the channel matrix. This is a function of the scattering environment and the antenna spacing [28]. Without loss of generality, the discussion in this thesis relates to spatial correlation at the transmitter. We consider a uniform linear array (ULA) placed on the x, y-plane as shown in Fig. 2-2. Now, the spatial correlation between two antenna elements, m

and m', i.e., $\mathbf{R}_{m,m'}$, where $m, m' \in 1, \ldots, M_T$ can be defined as [61]

$$\mathbf{R}_{m,m'} = \frac{\text{cov}[\mathbf{h}_m, \mathbf{h}_{m'}]}{\sqrt{\text{var}[\mathbf{h}_m]\text{var}[\mathbf{h}_{m'}]}}, \tag{2.13}$$

where $|\mathbf{R}_{m,m'}| = 0$ means the antenna elements are uncorrelated. In general, $0 \leq |\mathbf{R}_{m,m'}| \leq 1$ and $|\mathbf{R}_{m,m'}| = 1$ indicates that the channels at the two transmit antenna elements are linearly dependent on each other, or fully correlated [55] [61]. Spatial correlation between the antenna elements of the array increases when inter-element antenna spacing reduces [61]. Furthermore, the level of spatial correlation also depends upon the channel characteristics. For example, in a NLOS channel, with a wide angular spectra less spatial correlation occurs. This is because wide angular spectra provide more independent scattering among the transmitted/received signals.

In this thesis, in order to investigate the transmit spatial correlation in MU-MIMO systems and further to explore NOMA performance we employ the one-ring model for the spatial fading correlation matrices. The one-ring model is a geometry based stochastic model. In this model, the BS is usually elevated and not obstructed by local scatterers, however the receiving users are surrounded by local scatterers [63] [64] [65] as shown in Fig. 2-3 [65].

In a particular propagation environment, these models are generally parameterized in terms of the central azimuth angle, ϕ, angular spread, Δ, and inter-element spacing at the BS. We assume that the central azimuth angle, ϕ, and the LoS angle, $\bar{\phi}$, of any associated Rician channel are equal. The spatial correlation matrix of the user assuming an one-ring model is given by [63]

$$\mathbf{R}_{m,m'} = \frac{1}{2\Delta} \int_{-\Delta+\phi}^{\Delta+\phi} e^{-j2\pi d(m-m')\sin\alpha} d\alpha, \tag{2.14}$$

where $d(m - m')$ captures the inter-element spacing normalized by the carrier wavelength between the m-th and m'-th antenna elements. Indeed, we set $d = 0.25$ to maintain the correlation between the antennas. In chapter 6, we will investigate the effect of the spatially correlated environment on NOMA performance when consider-

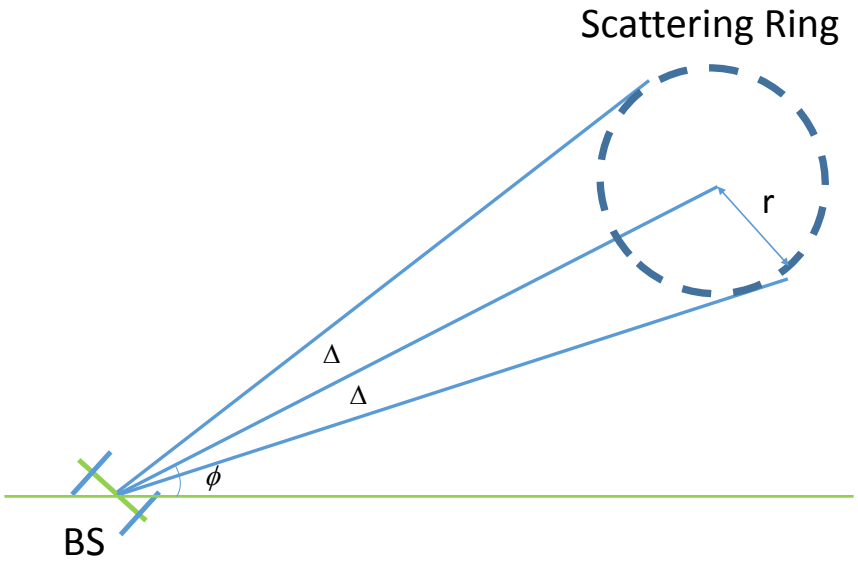


Figure 2-3: A user at central azimuth angle ϕ with a scattering ring of radius r generating an angular spread, Δ, on both sides of the central angle.

ing the one-ring model. Moreover, the effects of LoS and spatial correlation will be analyzed in Rayleigh and Rician fading channel models.

2.2 Multi-user MIMO (MU-MIMO) system

In a MU-MIMO system, a BS with multiple antennas communicates with multiple users, each with multiple receiving antennas as shown in Fig. 2-4. Without loss of generality, throughout this thesis we consider the multi-user downlink, where a BS with multiple antennas transmits multiple streams simultaneously to more than one user. These multiple data streams sharing the same time and frequency resources result in interference. This interference among the users' streams can be mitigated exploiting multiuser diversity, i.e., selecting the most diversified users with favorable channel conditions to improve the system performance [28]. In addition, the spatial dimensions can be leveraged to avoid interference, i.e., directing the signal energy towards the intended user and minimizing the interference to other users. This can be done in two ways, either by employing digital BF, which is known as precoding, or

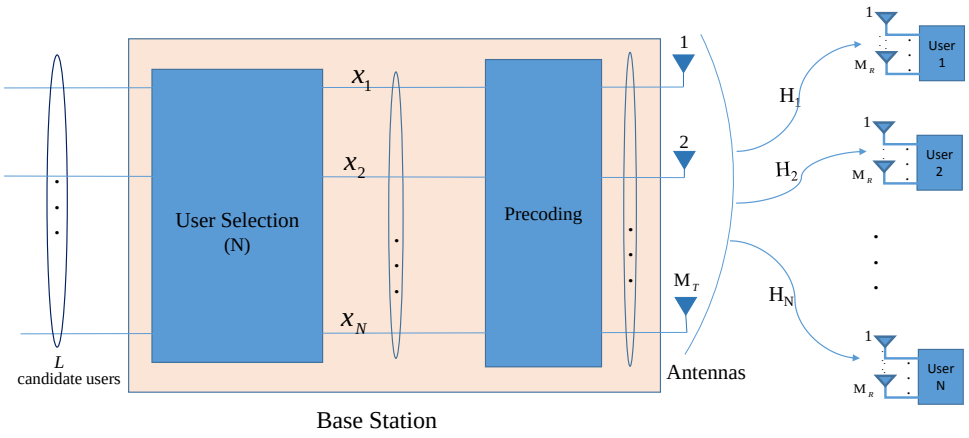


Figure 2-4: A MU-MIMO downlink system.

by employing analog BF simply known as BF, which are both described in Sec. 2.3.

We consider a MU-MIMO system, where a total of M_T transmitting antennas at the BS serves N UEs, each with M_R receiving antennas. The $NM_R \times M_T$ propagation channel matrix, $\mathbf{H}$, from the BS to the N users can be written as

$$\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \cdots, \mathbf{H}_N^T]^T. \tag{2.15}$$

Note that $\mathbf{H}_i$ is a $M_R \times M_T$ channel matrix for $i = 1, 2, \ldots, N$ and $\mathbf{H}_i = \sqrt{\beta_i}\mathbf{G}_i$. Furthermore, $\mathbf{H}$ is environment dependent, i.e., different propagation mechanisms influence the strength of the channels and the fading models differently [34]. Therefore, the channel matrix is modeled via the different fading channel models as discussed in Sec. 2.1. The strength of the channels was discussed in Sec. 2.1.1, hereinafter referred to as the link gain. Now, assuming $M_T \geq NM_R$, the downlink MIMO signal received by each user, $i \in 1, 2, \ldots, N$, can be expressed as [28]

$$\begin{aligned} \mathbf{r}_i &= \sqrt{\beta_i}\mathbf{G}_i\mathbf{x}_i + \mathbf{n}_i \\ &= \mathbf{H}_i\mathbf{x}_i + \mathbf{n}_i \end{aligned} \tag{2.16}$$

where β_i is the link gain, $\mathbf{x}_i$ is the transmitted vector signal of dimension M_T and $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{M_R})$ denotes the noise vector of dimension M_R.

2.2.1 Decomposition of the MIMO channel and receive processing design

In this thesis we consider both single antenna and multiple antennas at the receivers/users. In chapters 5 and 6 we consider a downlink system with single antennas at the receivers/UEs. Therefore, the channel from the multiple antennas at the BS to the users is represented as a channel vector, $\mathbf{h}_i$. Furthermore, in chapters 3 and 4 we consider multiple antennas at the receiver, where the receiver employs a linear combiner using the singular value decomposition (SVD) , converting the channel matrix, $\mathbf{H}_i$, to a channel vector $\mathbf{h}_i$ as given below. Hence, throughout the thesis each user has a channel, $\mathbf{h}_i$, which is a vector and represents either the actual vector channel or an effective channel vector created by receive combining.

Using the SVD, every matrix $\mathbf{H}_i \in \mathcal{C}^{M_R \times M_T}$ can be decomposed accordingly to its singular values [29]. The SVD of the channel is denoted as

$$\mathbf{H}_i = \mathbf{U}_i \mathbf{D}_i \mathbf{V}_i^H, \tag{2.17}$$

where $\mathbf{U}_i$ and $\mathbf{V}_i$ are $M_R \times M_R$ and $M_T \times M_T$ unitary matrices, respectively, and $\mathbf{D}_i = \text{diag}(d_1^{(i)}, d_2^{(i)}, \cdots, d_m^{(i)})$ contains the $m = \min(M_R, M_T)$ ordered singular values of $\mathbf{H}_i$ such that $d_1^{(i)} \geq d_2^{(i)} \geq \cdots \geq d_m^{(i)}$.

Denoting the first column of $\mathbf{U}_i$ by $\mathbf{u}_i$, the effective user channels after receive processing are then $\mathbf{h}_i = \mathbf{u}_i^H \mathbf{H}_i$, $i = 1, 2, \ldots, N$. This operation converts the matrix channel $\mathbf{H}_i$ of dimension $M_R \times M_T$ to a vector channel $\mathbf{h}_i$ of dimension $1 \times M_T$ [29]. Thus, the global channel from the BS to the N users in (2.15) can be re-written as

$$\mathbf{H} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_N^T]^T, \tag{2.18}$$

with dimension $N \times M_T$.

2.3 Precoding

In a closed loop MIMO system, the transmitter sends pilot signals to the receivers and the channel state information (CSI) is estimated by the channel estimators at the receivers. After CSI estimation, the receiver sends back the estimated channel information to the transmitter (CSIT). With the CSIT feedback, the transmitter designs the signals with appropriate weighting across all antennas such that the signal power is maximized at the receiver output [66], which is known as precoding. In a downlink MU-MIMO system, precoding is designed for separating data streams with as little interference as possible, i.e., the BS pre-filters signals to the intended users before sending them through the channel [67].

Precoding design for MIMO wireless systems is an active research area and there is a rapidly growing interest in precoding design techniques for MU-MIMO systems in 5G technology [68] [69] [13]. In MIMO systems, both non-linear precoding, such as DPC [70], Tomlinson-Harashima precoding (THP) [71], lattice-aided methods [72], and linear precoding techniques such as match filtering (MF), MMSE and ZF schemes [30] [31] [32] can be used. Non-linear methods have better performance compared to the linear precoding techniques, however, at the cost of higher implementation complexity. Moreover, linear precoders, such as MF and ZF, are shown to be near optimal with an increase in the number of antennas and users [33] [50] [48]. With careful selection of precoder and spatial separation between the users, inter-user interference (IUI) can be significantly reduced thereby supporting multiple users simultaneously. It is known that ZF outperforms MF in the high spectral efficiency region while MF is better in lower spectral efficiency region [73]. Finding the optimal precoding vector is a non-convex optimization problem [74].

In this thesis, we will consider ZF precoding to avoid the inter-stream interference which can be simply done by inverting the composite channel matrix of the users [74]. We will discuss in ZF and assumption in the next section.

2.3.1 Zero-forcing (ZF) precoding

ZF precoding is designed such that there is no inter-stream interference. After using receive processing, as discussed in sec. 2.2.1, $\mathbf{h}_i = \mathbf{u}_i^H \mathbf{H}_i$, $i = 1, 2, \ldots, N$, where each matrix channel $\mathbf{H}_i$ of dimension $M_R \times M_T$ is converted to a vector channel $\mathbf{h}_i$ of dimension $1 \times M_T$ [29]. Thus, the global channel $\mathbf{H}$ denotes the composite $N \times M_T$ downlink channel matrix from the BS to all N users.

The transmitter designs the ZF precoders, $[\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_N]$, by taking the pseudo-inverse of $\mathbf{H}$, namely $\mathbf{H}^\dagger$ [75]. Thus, the un-normalized precoding matrix, $\mathbf{W}_0$, is given by[5]

$$\mathbf{W}_0 = \mathbf{H}^\dagger = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} = [\mathbf{w}_{01}, \mathbf{w}_{02}, \cdots \mathbf{w}_{0N}], \tag{2.19}$$

so that the normalized precoder is

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N], \tag{2.20}$$

where

$$\mathbf{w}_i = \mathbf{w}_{0i} / ||\mathbf{w}_{0i}||. \tag{2.21}$$

Furthermore, the transmitter multiplies the data symbol for each user by a precoding vector $\mathbf{w}_i \in \mathcal{C}^{M_T \times 1}$. Thus, the transmitted signal is the linear function $x = \sum_{i=1}^{N} \mathbf{w}_i \mathbf{x}_i$. The resulting received signal for each user i can be written as [36]

$$\tilde{\mathbf{r}}_i = \mathbf{H}_i \mathbf{w}_i \mathbf{x}_i + \sum_{j=1,\, i \neq j}^{N} \mathbf{H}_i \mathbf{w}_j \mathbf{x}_j + \mathbf{n}_i, \tag{2.22}$$

[5]Note that the matrix $(\mathbf{H}\mathbf{H}^H)^{-1}$ must be a full rank matrix to be invertible.

followed by receive processing

$$\begin{aligned}\mathbf{r}_i &= \mathbf{u}_i^H \tilde{\mathbf{r}}_i \\ &= \mathbf{u}_i^H \mathbf{H}_i \mathbf{w}_i \mathbf{x}_i + \sum_{j=1,\, i\neq j}^{N} \mathbf{u}_i^H \mathbf{H}_i \mathbf{w}_j \mathbf{x}_j + \mathbf{u}_i^H \mathbf{n}_i, \\ &= \mathbf{h}_i \mathbf{w}_i \mathbf{x}_i + \sum_{j=1,\, i\neq j}^{N} \mathbf{h}_i \mathbf{w}_j \mathbf{x}_j + \mathbf{z}_i,\end{aligned} \tag{2.23}$$

where the second-term in (2.23) is the interference from other streams and $\mathbf{z}_i$ has the same statistics as the white Gaussian noise vector $\mathbf{n}_i$. The basic principle of ZF is to choose the precoders to avoid interference among user streams [76][77], i.e., the precoders are selected such that they satisfy

$$\mathbf{h}_i \mathbf{w}_j = 0 \text{ for } i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, N, i \neq j. \tag{2.24}$$

Therefore, the inter stream interference can be removed and (2.23) can be written as

$$\begin{aligned}\mathbf{r}_i &= \mathbf{h}_i \mathbf{w}_i \mathbf{x}_i + \mathbf{z}_i \\ &= \frac{\mathbf{x}_i}{||\mathbf{w}_{0i}||^2} + \mathbf{z}_i.\end{aligned} \tag{2.25}$$

However, ZF may suffer a power penalty due to the nulling. Note that, how much power can be allocated to the user or in a beam depends upon the channel condition. There is no power penalty if $\mathbf{H}$ is full rank and the Gram matrix ($\mathbf{H}\mathbf{H}^H$) is diagonal, i.e., the channels from BS to different users are orthogonal. However, when the Gram matrix becomes ill conditioned, i.e., some users are correlated or aligned to each other, the power penalty increases. In this situation the Gram matrix becomes large and power allocation matrix becomes small. Less energy is allocated to users and more energy is spent on nulling. Less power on user signals means low receive signal power at some users [30][67, p.17]. Therefore, in this thesis we consider $M_T \geq N$.

Let us briefly discuss about MF and MMSE precoding techniques. MF is the simplest linear precoding that aims to maximize the receive SNR at each user. MF is

more suitable in noise-limited scenarios and outperforms at low SNRs. However, at high SNRs, MF performance is limited by inter-user interference [67]. If we consider the entire SNR range we can find a trade-off between the signal strength and the interference reduction. In MMSE the main idea is that the noise variance is estimated at the receiver and fed back to the transmitter so that a better precoder can be designed for the entire SNR range. The MMSE solution is given as

$$\mathbf{W}_{\text{MMSE}} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I})^{-1}, \tag{2.26}$$

where α is the regularization parameter (please refer to [67, p. 19] for more information.). This (2.26) becomes when $\alpha = 0$, i.e., at very high SNRs, and converses to MF when $\alpha \to \infty$, i.e., at low SNRs.

Moreover, the precoder design for OMA systems and NOMA system is quite different. In TDMA system, global channel $\mathbf{H}$ considers all N users channels while designing the precoder. However, in NOMA systems, the precoder is designed based on the strong user channels in a cluster and singletons. Weak users channels are not considered which is elaborated in Sec. 2.4.4.

Signal to noise ratio (SNR)

The SNR of user i under ZF is

$$\text{SNR}_i = (E_x/\sigma^2)|\mathbf{h}_i\mathbf{w}_i|^2 = \rho|\mathbf{h}_i\mathbf{w}_i|^2 = \frac{\rho}{||\mathbf{w}_{0i}||^2}, \tag{2.27}$$

where $E_x = E[|x_i|]^2$ is the power of the data symbols, $x_1, x_2, \cdots, x_N$, sent to each user and $\rho = E_x/\sigma^2$, where σ^2 is the noise power at each receive antenna. Thus, the corresponding rate can be written as

$$R_i = \log_2(1 + \text{SNR}_i). \tag{2.28}$$

2.4 Multiple access techniques

Multiple access techniques can be classified as orthogonal or non-orthogonal. In this section we will describe both approaches. [6]

2.4.1 Orthogonal multiple access (OMA)

Orthogonal multiple access (OMA) means that signals from different users are orthogonal to each other. This can be achieved by time division multiple access (TDMA), frequency division multiple access (FDMA) or orthogonal frequency division multiple access (OFDMA). In these techniques, multiple users access and share the system resources simultaneously, but in a non-overlapping manner. In OMA, resources are allocated to the users in each cell such that there is no inter-user interference. Hence, low-complexity detection approaches can be implemented at the receiver to retrieve the users' signals [78]. Additionally, OMA can achieve good system level throughput performance in packet domain services using channel-aware time and frequency domain scheduling with simple single user detection [79]. However, the requirement of accommodating a dynamic set of different types of random active users in terms of bandwidth allocation in 5G cellular communication makes it unfavorable for future radio access networks [80]. Also in orthogonal systems, all associated BSs for coordinated multi-point (CoMP) transmission need to allocate the same channel to the cell-edge users and this restricts the channel being used by other users. This restriction in the reuse of the channel decreases the spectral efficiency with an increase in the number of cell edge users [81].

2.4.2 Non-Orthogonal multiple access (NOMA)

Unlike OMA, NOMA is an intra-cell multi-user multiplexing scheme that utilizes an additional new domain, i.e., the power domain, which has not been extensively utilized in previous 2G (TDMA), 3G (Code Division Multiple Access (CDMA)) or 4G (OFDMA) systems [7]. In 4G cellular communication OFDMA has been used which

[6] Precoding design is not significantly influenced by the multiple access techniques employed.

allows multiple access among several terminals while preserving frequency orthogonality. However, NOMA provides a multiple access technology utilizing power control in the frequency domain. Since several users' signals can be sent at the same time on the same frequency, higher numbers of users can be accommodated as well as increasing the overall cellular capacity.

NOMA is a promising technology for 5G, when used with advanced transmission (superposition coding (SC)) and reception (successive interference cancellation (SIC)) techniques[7]. The benefit of NOMA with SIC, along with the importance of resource sharing and user coordination in 5G networks has been the subject of many recent studies [6] [82] [83] [84]. The target of these studies is to investigate the feasibility and potential gains of replacing OMA with other multiple access schemes and to analyze the computational effort required for real usage. Moreover, some of the work [78] [85] [86] on uplink NOMA systems concluded that enhanced uplink performance can be achieved compared with OMA systems.

NOMA achieves a better trade-off between the total user throughput and user fairness with regard to the achievable throughput of the respective users [79]. In NOMA, specific functionalities such as random BF [10], and opportunistic BF [87], multi-user power allocation (non-equal power transmission among beams) [88], dynamic user multiplexing order [80], receiver design [89] and SIC with error propagation [6] [90] [91], are considered to maximize NOMA gains.

2.4.3 Superposition coding (SC) and successive interference cancellation (SIC)

SC is a coding technique that performs multiplexing of two or more users' signals in the power domain [92] [93]. For example, if there are two geographically distributed users one near the BS and the other far from the BS, then the BS transmits the superimposed signal of both near and far user, scaling both with different power levels. The probability of detection of each signal at the receiver increases with more

[7]SC and SIC techniques are described in Sec. 2.4.3.

distinct power levels of the transmitted superimposed signals. By SC the difference in signal strength, i.e., near-far effect at wireless communication of users, can be exploited to convey signals of multiple users over the same channel simultaneously.

SIC is used to decode the received superimposed signal, i.e., the receiver starts by decoding the message of the user having the strongest SNR, then continues in a descending order of the users' SNR [94] [95]. After detection of each user's signal, the receiver subtracts the detected signal from the received signal. This suppresses co-channel interference from the strong signal for relatively weak signals, and it is possible to successfully detect the signals of all users with high probability. Finally, SC and SIC make it possible to extend the capacity of a communication system without expanding the bandwidth [93].

2.4.4 NOMA-ZFBF

In this thesis, we consider NOMA-ZFBF systems in which highly correlated users channels share a single BF vector [40][96]. The basic NOMA-BF system in the multiuser system scenario has some beams supporting two or more highly correlated users in a cluster and the rest of the beams support singletons. Singletons are the users that cannot be grouped into highly correlated groups. We assume that the BS can support Ω users in a NOMA-ZFBF system[8] and all users are equipped with either single antenna or multiple antennas. As discussed in sec. 2.2.1, for a receiver with a single antenna, we have the channel $\mathbf{h}_i$. However, for a receiver with multiple antennas, the receiver uses a linear combiner so $\mathbf{h}_i$ is the effective channel, i.e., $\mathbf{h}_i = \mathbf{u}_i^H \mathbf{H}_i$.

To integrate NOMA with a multiuser system scenario, the Ω users are grouped into N_1 clusters with B highly correlated users in each cluster and N_2 singletons, i.e., $\Omega = N_1 B + N_2$. As shown in Fig. 2-5, we consider $B = 2$. ZFBF is used to separate the beams with clusters and the singletons. Within the clusters the strong and weak users are separated using NOMA power allocation and SIC [97].

[8]In TDMA systems each user is supported by a single beam, however in NOMA systems two or more users can be supported by a single beam. Thus, in NOMA systems, Ω represents the total number of supported users for transmission at a time and N is the total number of beams.

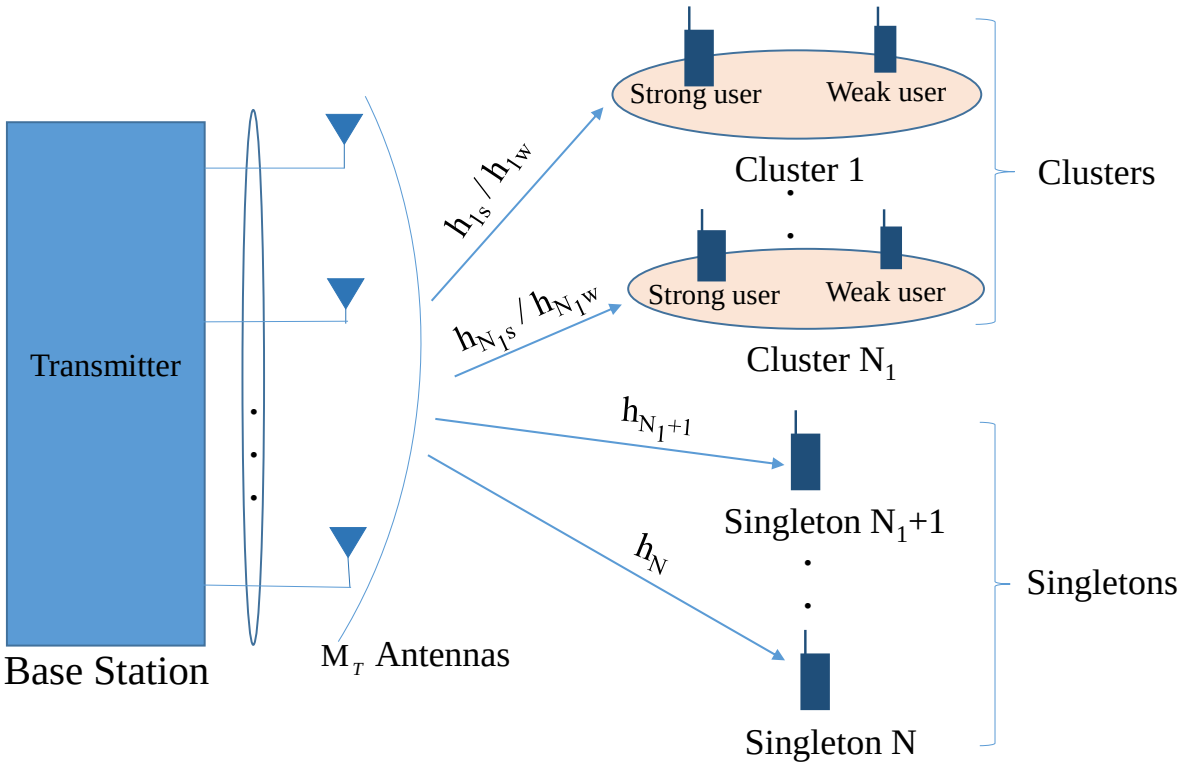


Figure 2-5: NOMA-BF downlink system with user selection and clustering algorithm.

Let $N = N_1 + N_2$, where $M_T \geq N$[9]. then the $N \times M_T$ channel matrix from the BS to the strong users in N_1 clusters and N_2 singletons is given by

$$\mathbf{H} = [\underbrace{\mathbf{h}_{1s}^T, \mathbf{h}_{2s}^T, \cdots, \mathbf{h}_{N_1 s}^T}_{\text{Clusters}}, \underbrace{\mathbf{h}_{N_1+1}^T, \cdots, \mathbf{h}_N^T}_{\text{Singletons}}]^T. \tag{2.29}$$

Throughout the thesis, we use subscripts s and w to denote the strong/near and weak/far users in the N_1 clusters, respectively. Note that (2.29) considers only the strong users and singletons channels, but weak users channels are not considered. Then, the transmitter designs the precoder based on $\mathbf{H}$ as in (2.19). This is different from (2.18), where global channel $\mathbf{H}$ considers all N users channels and the transmitter designs precoder based on (2.18) for OMA systems.

The BS sends the superimposed signal of both users multiplied with the precoder.

[9]Note that, NOMA system involves designing a single BF vector to support multiple users and there might be more number of users than number of transmitting antennas at the BS. However, we have not considered the overloading case where there are more users than number of antennas at the BS i.e, $M_T < N$.

The signals transmitted by the BS are given by

$$\mathbf{s} = \mathbf{W}\mathbf{x}, \tag{2.30}$$

where the $M_T \times 1$ vector $\mathbf{x}$ is given by

$$\mathbf{x} = \begin{bmatrix} \left(x_{1s}\alpha_1 + x_{1w}\sqrt{1-\alpha_1^2} \right. \\ \vdots \\ \left(x_{N_1 s}\alpha_{N_1} + x_{N_1 w}\sqrt{1-\alpha_{N_1}^2} \right. \\ x_{N_1+1} \\ \vdots \\ x_N \end{bmatrix}. \tag{2.31}$$

In clusters, the near user is served with less power as it has better channel conditions than the far user. Since ZF is only designed with the strong user's channels, $\mathbf{h}_{iw}\mathbf{w}_j \neq 0$ for $j \neq i$, the weak user suffers from interference from other clusters and from their strong user partner.

2.4.5 SNR/SINR of strong and weak user in clusters

The received signal for the near/strong user in the i^{th} cluster, where $i = 1, \ldots N_1$, is given by

$$r_{is} = \mathbf{h}_{is}\mathbf{w}_i\, x_i + \sum_{j=1,\, j\neq i}^{N} \mathbf{h}_{is}\mathbf{w}_j x_j + z_{is}, \tag{2.32}$$

where $x_1 \cdots x_{N_1}$ are the signals for the clusters given by $x_i = \left(x_{is}\alpha_i + x_{iw}\sqrt{1-\alpha_i^2}\right)$ and $x_{N_1+1} \cdots x_N$ are the signals for singletons. Further, x_{is} and x_{iw} are the strong user signal and weak user signal for cluster i, respectively. α_i and $\sqrt{1-\alpha_i^2}$ are the amplitude scalings and α_i^2 and $1-\alpha_i^2$ are the power fractions of x_{is} and x_{iw}, respectively.

Therefore, using $x_i = \left(x_{is}\alpha_i + x_{iw}\sqrt{1-\alpha_i^2}\right)$ for the i^{th} cluster, (2.32) can be written as

$$r_{is} = \mathbf{h}_{is}\mathbf{w}_i\Big(x_{is}\alpha_i + x_{iw}\sqrt{1-\alpha_i^2}\Big) + \sum_{j=1,\, j\neq i}^{N} \mathbf{h}_{is}\mathbf{w}_j\, x_j + z_{is}. \tag{2.33}$$

The second term of (2.33), $\sum_{j=1,j\neq i}^{N} \mathbf{h}_{is}\mathbf{w}_j x_j$ is the inter-stream interference from other beams and $\mathbf{h}_{is}\mathbf{w}_i\, x_{iw}\sqrt{1-\alpha_i^2}$ is the inter-user interference from the weak user in the same cluster. However, the ZFBF vector is generated according to the strong user's channel, $\mathbf{h}_{is}\mathbf{w}_j = 0$ for $1 \leq j \leq N$, $j \neq i$. Using this the inter stream interference from other beams can be removed. In addition, the strong user, employing perfect SIC, can remove the inter-user interference from the weak user. Thus, the received signal of the strong user can be written as

$$r_{is} = \mathbf{h}_{is}\mathbf{w}_i\, x_{is}\alpha_i + z_{is}. \tag{2.34}$$

Then, the received SNR can be written as

$$\mathrm{SNR}_{is} = (E_x/\sigma^2)|\mathbf{h}_{is}\mathbf{w}_i|^2\alpha_i = \rho|\mathbf{h}_{is}\mathbf{w}_i|^2\alpha_i. \tag{2.35}$$

Similarly, the received signal for a weak user in a cluster is given by

$$r_{is} = \mathbf{h}_{iw}\mathbf{w}_i\Big(x_{is}\alpha_i + x_{iw}\sqrt{1-\alpha_i^2}\Big) + \sum_{j=1,j\neq i}^{N} \mathbf{h}_{iw}\mathbf{w}_j x_j + z_{is}. \tag{2.36}$$

Since ZF is only performed with the strong user's channels, $\mathbf{h}_{iw}\mathbf{w}_j \neq 0$ for $j \neq i$, the weak user suffers from interference from other clusters and from their strong user partner. Hence, the SINR expression is more complex and is written as

$$\mathrm{SINR}_{iw} = \frac{\rho|\mathbf{h}_{iw}\mathbf{w}_i|^2(1-\alpha_i^2)}{\rho|\mathbf{h}_{iw}\mathbf{w}_i|^2\alpha_i^2 + \sum_{j=1,\, j\neq i}^{N} |\mathbf{h}_{iw}\mathbf{w}_j\, x_j|^2 + 1}. \tag{2.37}$$

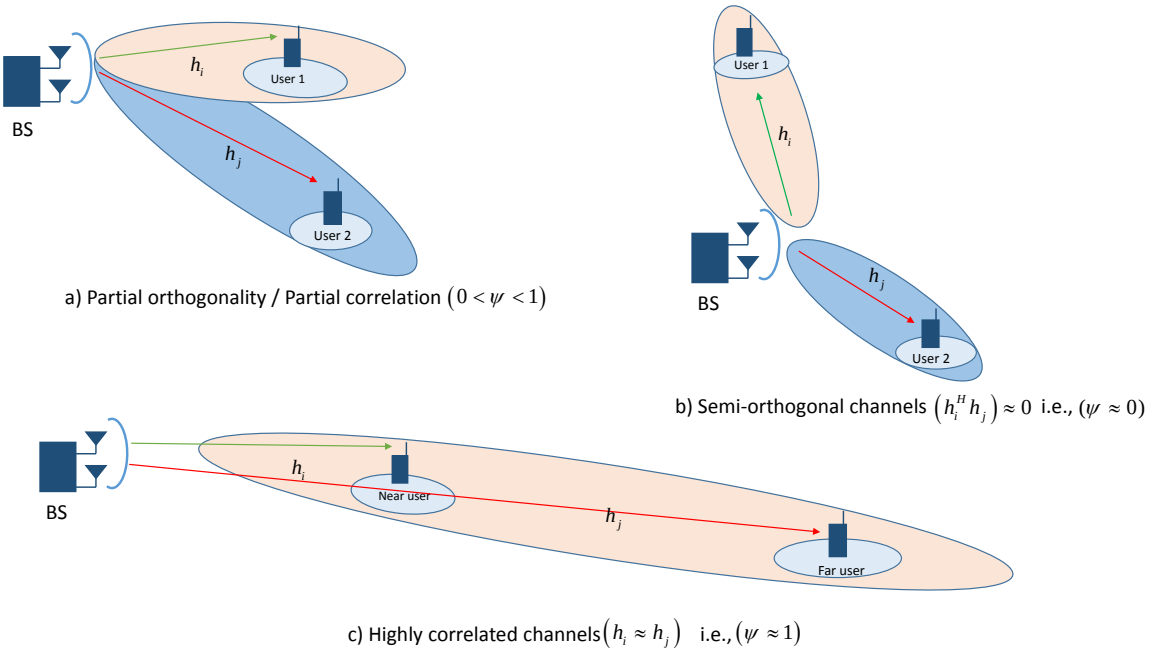


Figure 2-6: Illustration of correlation between the channels.

2.5 Correlation between the users

In a multiuser system, in order to exploit the multiuser diversity the BS can classify the users according to their spatial separability. This spatial separation can be quantified by the correlation between the spatial signatures of the users [98]. We can measure the instantaneous channel similarity of the users by correlating the channel vectors of each user as follows

$$\Psi = \frac{|\mathbf{h}_i \mathbf{h}_j^H|}{||\mathbf{h}_i|| ||\mathbf{h}_j||}. \tag{2.38}$$

Users with good spatial separation are known as semi-orthogonal users and those with little spatial separation are known as highly-correlated users. In order to better illustrate the concept of correlation between the user channels, we consider the case of two users as shown in Fig. 2-6 and describe the effects in terms of ZFBF.

Partial orthogonality / Partial correlation ($0 < \Psi < 1$)

As shown in Fig. 2-6 (a), in general the user channels are partially correlated with each other, i.e., their correlation satisfies $0 < \Psi < 1$. If the correlation is close to 0 then the users are semi-orthogonal and if the correlation is close to 1 then the users are highly correlated. The correlation value, Ψ, plays an important role while designing the beamforming weight vector or precoder as discussed in Sec. 2.3.1. Careful selection of users and designing their weight vectors accordingly can reduce the inter-stream interference which in turn increases the system performance.

Orthogonal channels ($\mathbf{h}_i^H \mathbf{h}_j \approx 0$)

As shown in Fig. 2-6 (b), as both users are in very different beam directions, the channels between the users may become orthogonal ($\Psi \approx 0$). Conventional ZFBF systems target the selection of users with semi-orthogonal channels as this reduces the noise inflation in the ZF approach. In this scenario, ZF performs well because they can easily satisfy the zero-interference condition, i.e., $\mathbf{h}_i \mathbf{w}_j = 0$ for $i \neq j$. A very effective example of this approach is the semiorthogonal user selection technique developed in [33]. However, targeting the selection of users with strong and orthogonal channels is not always effective in the scenarios where close proximity users with similar channels have to receive delay constrained traffic.

Highly correlated channels ($\mathbf{h}_i \approx \mathbf{h}_j$)

As shown in Fig. 2-6 (c), the channel between the users is highly correlated, i.e., $\mathbf{h}_i \approx \mathbf{h}_j$ or $\Psi \approx 1$, this is because the users are in the path of the same beam. An example could be a mmWave network deployed at a sports event with thousands of users, i.e., two or more users are in a highly directional mmWave beam. In this scenario, if all users in same beam are selected for transmission at the same time using ZF, each one of the users will have high interference from the other users which increases the noise inflation during channel inversion. Therefore, the primary issue in downlink precoding/beamforming for correlated users is how to balance the need for

a high received signal power for each user against the interference produced by the other user.

The NOMA-ZFBF technique can distinguish users via their power levels [40] [45] [96], where highly correlated users are grouped into clusters as discussed in Sec. 2.4.4. In [40], the ZFBF vector is generated based on the channel of the strongest user in a cluster to reduce the inter-stream interference and perform SIC correctly. If the BF vector was designed based on the weak user's channel, then there might be some errors due to the weak channel strength and it might make it difficult at the strong user to perform SIC due to the presence of inter-user interference.

However, the main concern about the above precoder design with the strong user's channel is: How does the weak user detect its information when the precoding channel (BF vector) $\mathbf{h}_j\mathbf{w}_i$ is not perfectly known? Only a few papers [27], [40], and [45] specify that NOMA with the above design requires highly correlated channels. Therefore, we have proposed a novel DZF technique in chapter 3 that can decorrelate channel vectors of highly correlated users. Further, in chapter 4 we have proposed a hybrid zero-forcing (HZF) technique combining both CZF and DZF that provides fairness against the joint scheduling of semi-orthogonal and highly correlated users.

Both, DZF and NOMA, handle the highly correlated user problem but in different ways. NOMA employs the same BF vector to highly correlated users and employs successive interference cancellation and power allocation to separate users [27] [40] to increasing the performance of the system. DZF uses simpler transceivers and instead of deliberately scheduling correlated users (as in NOMA), the scheme enables ZF to operate successfully in all scenarios, even in the presence of highly correlated users. In chapters 5 and 6, we investigate the NOMA-ZFBF system.

2.6 Simulation

We use Matlab to create the simulation results presented in this thesis. Unless otherwise stated, we consider a downlink MIMO system in either I.I.D. Rayleigh channel or correlated Rician channel environment. These can be simply constructed as

given in (2.4) and (2.12), respectively. For example, a simple Rayleigh channel with channel gain, β, is generated in Matlab using $h = \sqrt{\beta/2} \times (\text{randn}(M_R, M_T) + i \times \text{randn}(M_R, M_T))$. We assume that the BS knows the CSI of all users. Moreover, our simulation approach excludes the coding and modulation scheme in its performance evaluation, as our performance measures of SNR/SINRs and rates are independent of modulation. Hence, our results are valid for any modulation techniques. The benefit of using SNR/SINRs is that it enables us to obtain performance analysis of a wireless system without having to go into the details of transceiver design, or the coding and modulation schemes employed.

In Chaps. 3 and 4, we use $M_R = 2$ antennas at each receiver and for NOMA systems in Chaps. 5 and 6 we use a single antenna at each user, i.e, $M_R = 1$. The noise power is set to $\sigma^2 = 1$ without loss of generality. The channel gains of each user, i.e. strong, weak and singletons are specified in the respective chapters as are the other required parameters. In some cases the number of users is determined by the user selection algorithm, which in other cases the number of users is pre-determined. We run at least 100,000 simulations for each CDF and expected value point in the SNR and rate curves presented.

2.7 Summary

This chapter has provided the relevant background information required to understand the assumptions, models and techniques that are applied throughout the thesis. Specifically, we discussed the different channel models with and without spatial correlation, MU-MIMO systems with precoding, and correlation among the users.

Chapter 3

Decorrelating Zero-Forcing for Correlated MU-MIMO Channels

3.1 Introduction

In section 2.5 we showed that in MU-MIMO systems constructing a semi-orthogonal user group and using ZFBF increases the system performance. This is due to the ZFBF design process, since there is minimal loss in channel gains during channel-inversion.

Conventional zero-forcing (CZF) is a popular downlink processing scheme for MU-MIMO systems [30] [60], but performance degrades when the users' channels are correlated [37] [38]. As a result, most approaches to ZF in MU-MIMO systems target the selection of users with strong, uncorrelated channels as this reduces the noise inflation in the ZF approach. A very effective example of this approach is the semi-orthogonal user selection technique developed in [33]. Of course, the scheduler may have other constraints, such as fairness, and so it is useful to consider variations of ZF which are robust to the selection of users with similar channels, for example, when users are located in close proximity.

Emerging papers in 5G are discussing dense user environments and ultra-high connectivity [40] [41] [27] where close-proximity users and their associated highly correlated channel vectors are becoming more prevalent. Moreover, industrial field

trials have found that the linear processing technique is poor at separating users with similar channels [43] [44]. Hence, in these scenarios, it is reasonable to implement techniques such as spatial separation of users via massive BS arrays [41] or NOMA [40] [27].

In this chapter, we first demonstrate that highly correlated users are extremely likely in 5G wireless channels and are common even in simple Rayleigh fading channels for small numbers of transmitting antennas. Second, we propose a novel technique where users with MIMO arrays employ receive processing to decorrelate their effective channels. Unlike other techniques [41] - [44], the proposed algorithm is a variant of CZF with only a minimal change in system design and very little additional complexity. We split the set of users selected for service into two groups: the first group consists of any pairs of highly correlated users and the second group contains users which are not highly correlated with the others. The paired users are labelled "near" or "far" corresponding to a stronger or weaker channel. The proposed decorrelating ZF (DZF) design is similar to CZF in that for far users and unpaired users, the BS performs ZF and the users employ the leading left singular vectors for receive processing [33]. Unlike CZF, the near users employ the second strongest left singular vector for receive processing. As near and far users have highly correlated channels and the first two singular vectors are orthogonal, the proposed design aims for decorrelation (orthogonalization) of the near and far user effective channels, hence reducing noise inflation [23]. Of course, this noise reduction comes at the expense of a power loss for the near user. However, analytical and simulation results show that the proposed precoder design can provide performance gains over CZF when correlated users are present. Moreover, we show that our proposed technique achieves better fairness than CZF according to Jain's fairness measure [99].

3.2 Multi-user MIMO system model

Consider a MU-MIMO system in which the BS is equipped with M_T transmitting antennas. The BS supports N UEs each with M_R receiving antennas. Assuming each

user experiences uncorrelated Rayleigh fading, the downlink channel matrix, $\mathbf{H}_i$, for user i contains entries which are $\mathcal{CN}(0, \beta_i)$, where β_i is the link gain for user i. The SVD of the channel is denoted

$$\mathbf{H}_i = \mathbf{U}_i \mathbf{D}_i \mathbf{V}_i^H, \tag{3.1}$$

where $\mathbf{U}_i$ and $\mathbf{V}_i$ are $M_R \times M_R$ and $M_T \times M_T$ unitary matrices, respectively, and

$$\mathbf{D}_i = \text{diag}(d_1^{(i)}, d_2^{(i)}, \cdots, d_m^{(i)}), \tag{3.2}$$

contains the $m = \min(M_R, M_T)$ ordered singular values of $\mathbf{H}_i$ such that $d_1^{(i)} \geq d_2^{(i)} \geq \cdots \geq d_m^{(i)}$.

3.2.1 Conventional precoder design

A conventional approach to delivering one stream per user is to use the dominant left singular vector of the channel at each user and plain ZF at the BS. We refer to this technique as CZF. Denoting the first column of $\mathbf{U}_i$ by $\mathbf{u}_i$, the effective user channels are then $\mathbf{h}_i = \mathbf{u}_i^H \mathbf{H}_i$, $i = 1, \ldots, N$ and the global channel from the BS to the N users is

$$\mathbf{H} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_N^T]^T. \tag{3.3}$$

The transmitter performs ZF based on $\mathbf{H}$, assuming $M_T \geq N$. The un-normalized precoding matrix, $\mathbf{W}_0$, is given by

$$\mathbf{W}_0 = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} = [\mathbf{w}_{01}, \mathbf{w}_{02}, \cdots \mathbf{w}_{0N}], \tag{3.4}$$

so that the normalized precoder is

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N], \tag{3.5}$$

where $\mathbf{w}_i = \mathbf{w}_{0i} / ||\mathbf{w}_{0i}||$.

3.2.2 Proposed precoder design (DZF)

Here we design DZF, a simple alternative to CZF that is resilient to the presence of correlated users and outperforms CZF when correlated users are present. During precoder design, the BS obtains channel state information (CSI) feedback from users and decides if the users are highly correlated and further sets the precoders. Moreover, the BS sends 1-bit information to each user informing them to use the 1st or 2nd singular vector as the precoder. This has minimal overhead compared to CZF.

Assume that the $N = 2N_1 + N_2$ users are grouped into N_1 pairs (each containing a highly correlated near/strong (s) and weak/far (w) user) and N_2 singletons (users that can't be grouped into a highly correlated pair). As in [40], a predefined correlation threshold can be used for pair selection. Accordingly we relabel the channel matrices so they are ordered as $\mathbf{H}_{1s}, \cdots, \mathbf{H}_{N_1 s}, \mathbf{H}_{1w}, \cdots, \mathbf{H}_{N_1 w}, \mathbf{H}_{N_1+1}, \cdots, \mathbf{H}_N$ and each user pair, i, contains the highly correlated channels $(\mathbf{H}_{is}, \mathbf{H}_w)$.

The proposed design is similar to CZF in that the BS performs traditional ZF and the singletons and far users employ the leading left singular vectors for processing. In contrast, the near users employ the second strongest left singular vector which decorrelates their channel with the paired far user at the expense of a loss in power. Hence, the users perform receive processing to create equivalent channels as follows:

$$\mathbf{h}_{is} = \mathbf{u}_{is}^H \mathbf{H}_{is}, \;\; \mathbf{u}_{is} = \text{col. 2 of } \mathbf{U}_{is}, \;\; \mathbf{H}_{is} = \mathbf{U}_{is}\mathbf{D}_{is}\mathbf{V}_{is}^H, \tag{3.6}$$

$$\mathbf{h}_w = \mathbf{u}_w^H \mathbf{H}_w, \;\; \mathbf{u}_{iw} = \text{col. 1 of } \mathbf{U}_{iw}, \;\; \mathbf{H}_{iw} = \mathbf{U}_{iw}\mathbf{D}_{iw}\mathbf{V}_{iw}^H, \tag{3.7}$$

for $i = 1, \ldots, N_1$ and

$$\mathbf{h}_i = \mathbf{u}_i^H \mathbf{H}_i, \;\; \mathbf{u}_i = \text{col. 1 of } \mathbf{U}_i, \;\; \mathbf{H}_i = \mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^H, \tag{3.8}$$

for $i = N_1 + 1, \ldots, N$. Now, the global equivalent channel matrix, $\widetilde{\mathbf{H}}$, considering all users is given as

$$\widetilde{\mathbf{H}} = [\mathbf{h}_{1s}^T, \cdots, \mathbf{h}_{N_1 s}^T, \mathbf{h}_{1w}^T, \cdots, \mathbf{h}_{N_1 w}^T, \mathbf{h}_{N_1+1}^T, \cdots, \mathbf{h}_N^T]^T. \tag{3.9}$$

Then, the transmitter performs ZF using $\widetilde{\mathbf{W}}_0 = \widetilde{\mathbf{H}}^H(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H)^{-1}$, and the normalized precoder, $\widetilde{\mathbf{W}} = [\widetilde{\mathbf{w}}_1, \widetilde{\mathbf{w}}_2, \cdots \widetilde{\mathbf{w}}_N]$, is calculated similarly, where $\widetilde{\mathbf{w}}_i = \widetilde{\mathbf{w}}_{0i}/||\widetilde{\mathbf{w}}_{0i}||$.

3.2.3 Performance metrics

The SNR of ZF is well known so that the SNR[1] of user i under CZF is [40]

$$\mathrm{SNR}_i = (E_x/\sigma^2)/||\mathbf{w}_{0i}||^2 = \rho/||\mathbf{w}_{0i}||^2 \tag{3.10}$$

and under DZF is

$$\widetilde{\mathrm{SNR}}_i = (E_x/\sigma^2)/||\widetilde{\mathbf{w}}_{0i}||^2 = \rho/||\widetilde{\mathbf{w}}_{0i}||^2. \tag{3.11}$$

Note that $E_x = E[|x_i|]^2$ is the signal strength and $\rho = E_x/\sigma^2$, where σ^2 is the noise power at each receive antenna. The corresponding rates are

$$R_i = \log_2(1 + \mathrm{SNR}_i) \tag{3.12}$$

and

$$\widetilde{R}_i = \log_2(1 + \widetilde{\mathrm{SNR}}_i). \tag{3.13}$$

In the proposed technique, unlike CZF, the near and far users employ orthogonal receive combiners which tend to decorrelate the effective channels. There are links between DZF and NOMA, but the focus is quite different. NOMA focuses on increased capacity by using the same BF vector to highly correlated users and using successive interference cancellation and power allocation to separate users [40] [27]. DZF uses simpler transceivers and instead of deliberately scheduling correlated users (as in NOMA), the scheme enables ZF to operate successfully in all scenarios, even in the presence of highly correlated users.

[1]Note that when the transmitter has perfect CSI, ZF completely eliminates multi-user interference. However, if there is any imperfection in the channel knowledge, there will be some multiuser interference [100] [101].

3.3 Two user case

The benefits of our DZF precoder design can be seen using a correlated pair with channels

$$\mathbf{H}_s = \mathbf{U}_s \mathbf{D}_s \mathbf{V}_s^H, \tag{3.14}$$

and

$$\mathbf{H}_w = \mathbf{U}_w \mathbf{D}_w \mathbf{V}_w^H, \tag{3.15}$$

where$||\mathbf{H}_s|| \geq ||\mathbf{H}_w||$.

3.3.1 CZF

In CZF, user s employs receive combiner $\mathbf{u}_s$ which is the first column of $\mathbf{U}_s$. Similarly, user w employs the combiner $\mathbf{u}_w$ which is the first column of $\mathbf{U}_w$. Therefore, the effective channels can be written as

$$\mathbf{H}_{eff}^{(s)} = \mathbf{u}_s^H \mathbf{H}_s = \sqrt{\lambda_1^{(s)}} \mathbf{v}_s^H, \tag{3.16}$$

where $\mathbf{v}_s$ is the first column of $\mathbf{V}_s$ and $\lambda_1^{(s)}$ is the maximum eigenvalue of $\mathbf{H}_s \mathbf{H}_s^H$. Similarly

$$\mathbf{H}_{eff}^{(w)} = \mathbf{u}_w^H \mathbf{H}_w = \sqrt{\lambda_1^{(w)}} \mathbf{v}_w^H, \tag{3.17}$$

where $\mathbf{v}_w$ is the first column of $\mathbf{V}_w$ and $\lambda_1^{(w)}$ is the maximum eigenvalue of $\mathbf{H}_w \mathbf{H}_w^H$. Now, the SNR can be computed as

$$\text{SNR}_i = \rho \left(\left[\left(\mathbf{H}_{eff} \mathbf{H}_{eff}^H \right)^{-1} \right]_{ii} \right)^{-1}, \tag{3.18}$$

where $\mathbf{H}_{eff} = \left[\mathbf{H}_{eff}^{(s)T} \mathbf{H}_{eff}^{(w)T} \right]^T$. Further, $\mathbf{H}_{eff} \mathbf{H}_{eff}^H$ can be computed as

$$\mathbf{H}_{eff} \mathbf{H}_{eff}^H = \begin{bmatrix} \lambda_1^{(s)} & \sqrt{\lambda_1^{(s)} \lambda_1^{(w)}} \mathbf{v}_s^H \mathbf{v}_w \\ \sqrt{\lambda_1^{(s)} \lambda_1^{(w)}} \mathbf{v}_w^H \mathbf{v}_s & \lambda_1^{(w)} \end{bmatrix}. \tag{3.19}$$

Defining $\Gamma_c = \mathbf{v}_s^H \mathbf{v}_w$ and substituting (3.19) in (3.18) gives

$$\mathrm{SNR}_s = \rho\lambda_1^{(s)}(1 - |\Gamma_c|^2) \tag{3.20}$$

and

$$\mathrm{SNR}_w = \rho\lambda_1^{(w)}(1 - |\Gamma_c|^2). \tag{3.21}$$

Perfectly orthogonal case

When two singular vectors are perfectly orthogonal, i.e., $\Gamma_c = 0$, then the SNRs can be written as

$$\mathrm{SNR}_s = \rho\lambda_1^{(s)} \tag{3.22}$$

and

$$\mathrm{SNR}_w = \rho\lambda_1^{(w)}. \tag{3.23}$$

Perfectly correlated case

When both channels are perfectly correlated, $\mathbf{H}_n \propto \mathbf{H}_w$, then $\Gamma_c = 1$ and $\mathrm{SNR}_s = \mathrm{SNR}_w = 0$. Hence the SNRs decay quadratically with $|\Gamma_c|$ from $\rho\lambda_1^{(s)}$ and $\rho\lambda_1^{(w)}$ to zero as $|\Gamma_c|$ increases from zero to one, i.e, as the correlation increases.

3.3.2 DZF

In the proposed DZF design, unlike CZF, user s uses the combiner $\mathbf{u}_s$ which is the second column of $\mathbf{U}_s$. However, user w uses the combiner as in CZF, which is $\mathbf{u}_w$, the first column of $\mathbf{U}_w$. The effective channels can be written as

$$\mathbf{H}_{eff}^{(s)} = \mathbf{u}_s^H \mathbf{H}_s = \sqrt{\lambda_2^{(s)}}\mathbf{v}_s^H, \tag{3.24}$$

where $\lambda_2^{(s)}$ is the second largest eigenvalue of $\mathbf{H}_w\mathbf{H}_s^H$ and $\mathbf{v}_s$ is the second column of $\mathbf{V}_s$. Similarly,

$$\mathbf{H}_{eff}^{(w)} = \mathbf{u}_w^H\mathbf{H}_w = \sqrt{\lambda_1^{(w)}}\mathbf{v}_w^H, \tag{3.25}$$

where $\lambda_1^{(w)}$ is the maximum eigenvalue of $\mathbf{H}_w\mathbf{H}_f^H$ and $\mathbf{v}_w$ is the first column of $\mathbf{V}_w$. Now, $\mathbf{H}_{eff} = \left[\mathbf{H}_{eff}^{(s)T}\mathbf{H}_{eff}^{(w)T}\right]^T$ so that

$$\mathbf{H}_{eff}\mathbf{H}_{eff}^H = \begin{bmatrix} \lambda_2^{(s)} & \sqrt{\lambda_1^{(w)}\lambda_2^{(s)}}\mathbf{v}_s^H\mathbf{v}_w \\ \sqrt{\lambda_1^{(w)}\lambda_2^{(s)}}\mathbf{v}_w^H\mathbf{v}_s & \lambda_1^{(w)} \end{bmatrix}. \tag{3.26}$$

Defining $\Gamma_D = \mathbf{v}_s^H\mathbf{v}_w$ and substituting (3.26) in (3.18) gives the SNRs as

$$\widetilde{\mathrm{SNR}}_s = \rho\lambda_2^{(s)}(1 - |\Gamma_D|^2) \tag{3.27}$$

and

$$\widetilde{\mathrm{SNR}}_w = \rho\lambda_1^{(w)}(1 - |\Gamma_D|^2). \tag{3.28}$$

Again, the SNRs vary quadratically with a measure of similarity between eigenvectors, in this case, Γ_D.

Perfectly correlated case

When the two channels are perfectly correlated then $\mathbf{v}_s$ and $\mathbf{v}_w$ are orthogonal and $\Gamma_D = 0$ giving $\widetilde{\mathrm{SNR}}_s = \rho\lambda_2^{(s)}$ and $\widetilde{\mathrm{SNR}}_f = \rho\lambda_1^{(w)}$. Again, we have a quadratic drop in SNR as $|\mathbf{v}_s^H\mathbf{v}_w|$ varies. However, this time the pattern is reversed and SNR drops as correlation decreases.

Note that, the second singular vector has more power than the third vector and so on, so the second singular vector should always be used to ensure a reasonable sum-rate for the DZF.

3.4 Simulation results

First, we investigate the likelihood of very similar channels. If $\mathbf{h}_{1i}$, $\mathbf{h}_{1j}$ represent the first rows of the channel matrices $\mathbf{H}_i$, $\mathbf{H}_j$ for users i, j, then we measure the instantaneous row similarity by the normalized cross product[2],

$$C = |\mathbf{h}_{1i}\mathbf{h}_{1j}^H|(||\mathbf{h}_{1i}|| \times ||\mathbf{h}_{1j}||)^{-1}. \tag{3.29}$$

Note that $C = 0$ or $C = 1$ represent orthogonal or perfectly aligned channel vectors respectively. For two independent iid Rayleigh channels, it is known that [100]

$$\sqrt{1 - C^2} \sim \text{Beta}(2(M_T - 1), 1). \tag{3.30}$$

Using the cumulative distribution function (CDF) of the Beta distribution we obtain

$$P(C > c_0) = (1 - c_0^2)^{M_T - 1} \tag{3.31}$$

giving a closed form result for the probability that the instantaneous row similarity exceeds a given threshold, c_0. For more complex channel models, we use simulation and focus on the 3GPP channel model in [24]. This is a type of Saleh-Valenzuela (SV) channel model with 20 clusters and 20 sub-paths per cluster parameterized for a uniform linear array in an urban macro (UMa) non-line-of-sight (NLOS) environment (using parameters in [102, Table I]). In addition, we consider the spatial consistency procedure where user channels vary smoothly over a grid of size $d_{\text{corr}} \times d_{\text{corr}}$, where $d_{\text{corr}} = 50$m is the correlation distance for UMa NLOS channels. Analytical and simulated results for $P(C > c_0)$ are given in Table 3.1. In the table, 0.7 and 0.9 are the correlations between the users and the percentage exceedances for the correlation value are tabulated for different numbers of transmitting antennas i.e., $M_T = 2$ and 8. It is clear that similar channels tend to occur for independent iid Rayleigh channels only for small $M_T = 2$. In contrast, the SV models experience similar channels quite

[2]Although, in this case C only measures the row similarity, it can still be used to measure channel similarity by measuring C for each row.

Table 3.1: Percentage exceedances for channel similarity

Channel Model	$\% > 0.7$		$\% > 0.9$	
	M_T=2	M_T=8	M_T=2	M_T=8
iid Rayleigh	51	0.90	19	0.0009
SV independent ($> 50m$)	82.89	50.90	55.04	9.13
SV dependent ($d = 25m$)	90.38	86.21	71.18	35.49
SV dependent ($d = 5m$)	99.68	99.98	97.98	99

commonly, even for independent users. If spatial consistency is considered, then channel similarity becomes prevalent as the spatial separation of users is decreased.

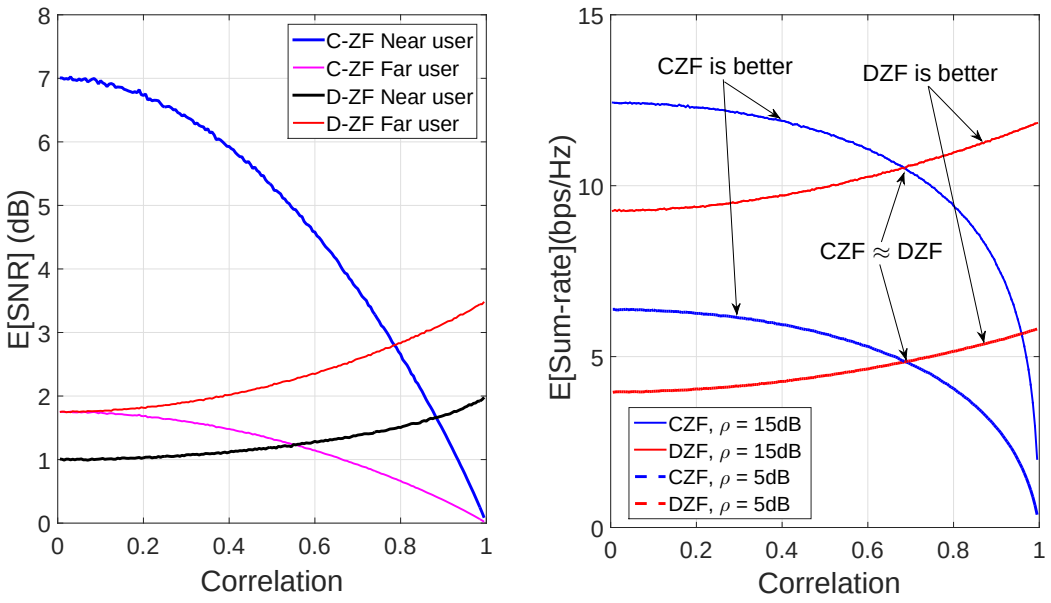
Next, we consider a system with $M_T = 2$ antennas at the BS and $M_R = 2$ antennas at all users unless otherwise specified. To construct the channels of highly correlated users, we first generate an uncorrelated Rayleigh fading channel matrix for the near user, $\mathbf{H}_s$, with elements $\mathcal{CN}(0, \beta_s = 4)$, we then generate a correlated channel,

$$\mathbf{H}_w = \sqrt{\beta_w/\beta_s}(\Psi\mathbf{H}_s + \sqrt{1-\Psi^2}\mathbf{E}_s) \tag{3.32}$$

with elements $\mathcal{CN}(0, \beta_w = 1)$, where Ψ is the correlation between the near and far user and the elements of $\mathbf{E}_s$ are $\mathcal{CN}(0, \beta_s)$. This allows us to control correlation through Ψ. The remaining N_2 unpaired users have independent Rayleigh channel matrices with elements $\mathcal{CN}(0, \beta_i = 1)$.

Figures 3-1a and 3-1b give mean SNR vs correlation and mean sum-rate vs correlation results for the case of two correlated users. As shown by the analysis in Sec. 3.3, CZF experiences a decay in both SNR and sum-rate when the correlation between the users increases. However, with DZF, the trend is reversed and SNR and sum-rate increase as correlation increases. Moreover, in Fig. 3-1b we observe that the crossover point, where the sum-rate of DZF exceeds CZF is at $\Psi \approx 0.7$. Further, as expected, the crossover point remains unaffected for both $\rho = 5$ dB and 15 dB since ρ simply scales the SNR.

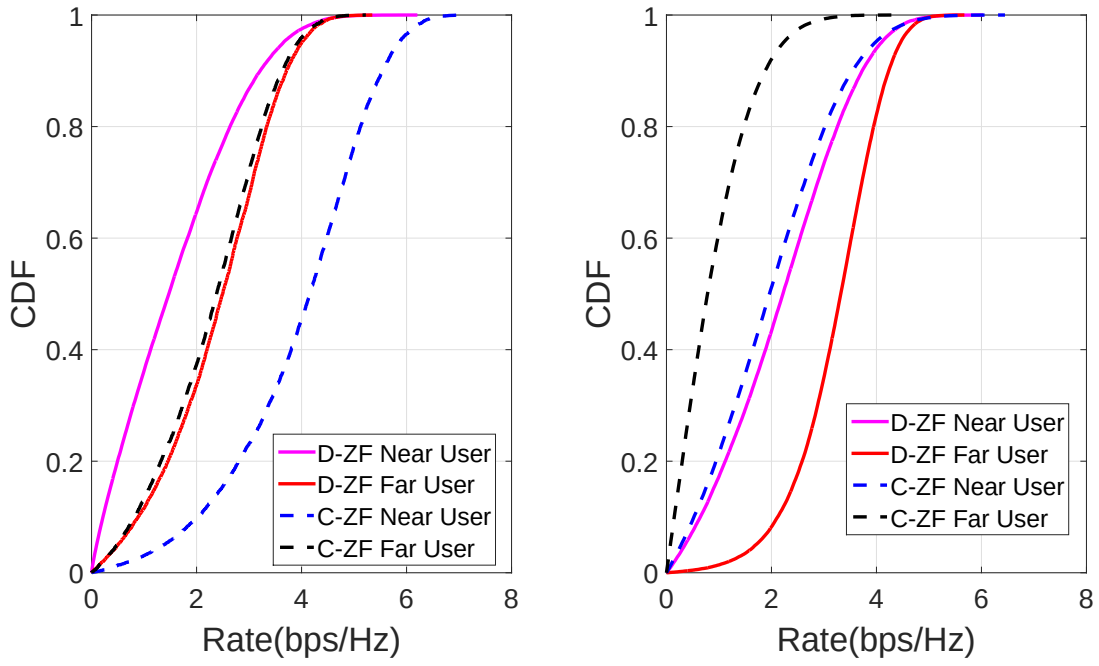
For the same scenario, in Fig. 3-2a and Fig. 3-2b, we plot rate CDFs for the near and far users at $\Psi = 0.2$ and $\Psi = 0.9$. These CDFs are obtained through simulation. The superiority of CZF at low correlation ($\Psi = 0.2$) is clear in Fig. 3-2a while for high



(a) SNR vs. correlation. $\rho = 5$ dB.

(b) Sum-rate vs. correlation

Figure 3-1: SNR and sum-rate comparison of DZF and CZF. $N = 2$, $\beta_s = 4$ and $\beta_w = 1$.



(a) $\Psi = 0.2$

(b) $\Psi = 0.9$

Figure 3-2: Rate CDF comparison of 2 users with DZF and CZF. $\rho = 5$ dB, $M_T = M_R = 2$, $N = 2$, $\beta_s = 4$ and $\beta_w = 1$.

correlation ($\Psi = 0.9$), Fig. 3-2b shows that the performance of near and far users is better for DZF.

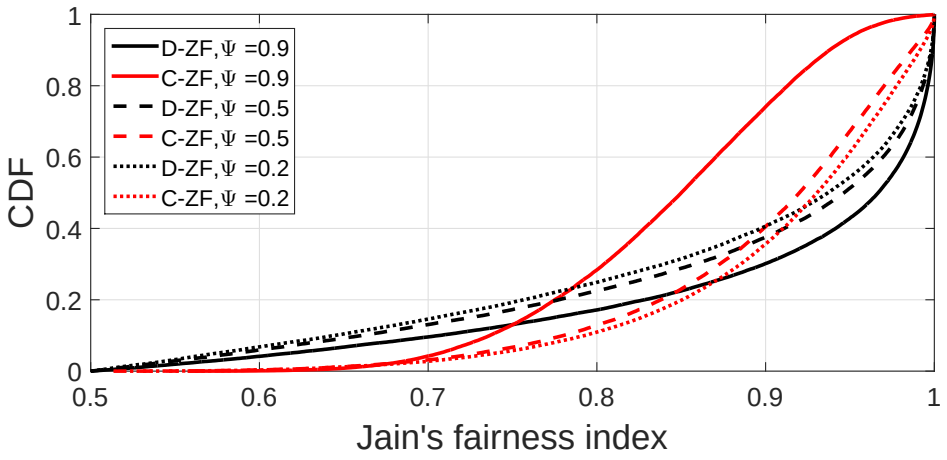


Figure 3-3: CDF of Jain's fairness index. $\rho = 5$ dB, $M_T = M_R = 2$, $N = 2$, $\beta_s = 4$ and $\beta_w = 1$.

To study the fairness of CZF and DZF, in Fig. 3-3 we plot the CDFs of Jain's fairness index [99] for the near and far user rates for the scenario in Fig. 3-2 with three correlation levels. The Jain's fairness index for CZF and DZF is calculated as

$$J = \frac{\left[\sum_{i=1}^{N} R_i\right]^2}{N \times \sum_{i=1}^{N} R_i^2} \text{ and } \widetilde{J} = \frac{\left[\sum_{i=1}^{N} \widetilde{R}_i\right]^2}{N \times \sum_{i=1}^{N} \widetilde{R}_i^2}, \tag{3.33}$$

respectively, where $N = 2$ and $i = 1, 2$ corresponds to near and far users. Note that 0.5 and 1 are the minimum and maximum fairness levels for this scenario. It is clear that DZF becomes fairer as Ψ increases while CZF becomes less fair. For high correlation ($\Psi = 0.9$), DZF is fairer than CZF for approximately 90% of the time. This is due to the built in fairness of DZF which involves a power penalty for the near user. For $\Psi = 0.9$, the cross-over at low fairness levels is due to the fact that sometimes the near user's power penalty reduces their rate relative to the far user.

In Fig. 3-4, we demonstrate the sum-rate dependence on the channel powers. We observe that increasing the strong user power from $\beta_s = 1$ to 4 also increases the overall sum-rate. Furthermore, increasing the channel gain power difference between

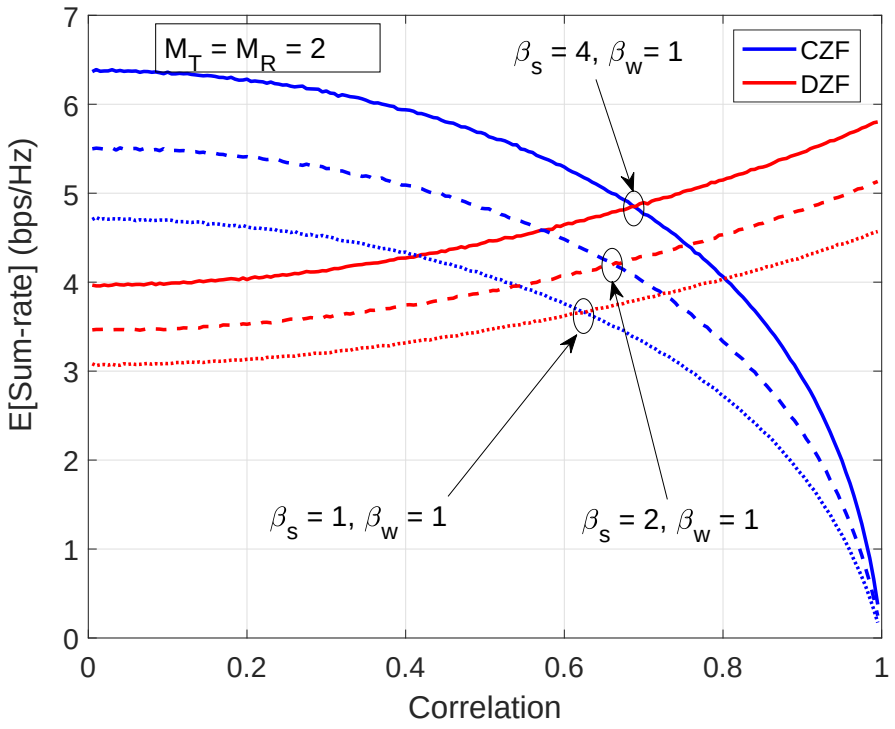


Figure 3-4: Sum-rate versus correlation. $\rho = 5$ dB and $N = 2$.

the strong and weak user shifts the crossover point to the right. This is because as β_s increases, the power penalty is more significant i.e., scaling the channel powers by some factor has more effect on larger powers, and therefore a higher correlation is needed to counter this effect.

Finally, in Fig. 3-5, we demonstrate the effect of varying the numbers of transmitting antennas and participating users on the sum-rate. It is obvious that as the number of transmitting antennas and users increases, the sum-rate of the system also increases. As in Fig. 3-4, as the sum-rate increases the cross-over point slightly shifts to the right as the effect of a single correlated pair of users is proportionally less in a larger global channel. However, with the increase in the number of correlated pairs the cross-over point shifts to left and the sum-rate of CZF decreases rapidly.

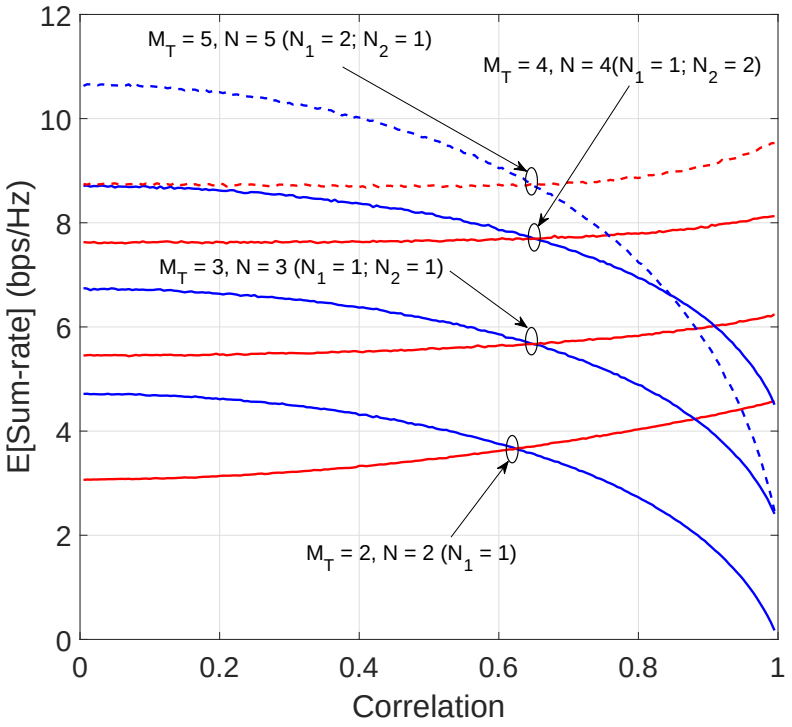


Figure 3-5: Sum-rate versus correlation with different M_T and N. $\rho = 5$ dB, $M_R = 2$, $\beta_s = 1$ and $\beta_w = 1$.

3.5 Conclusion

In this chapter, we have shown that DZF is an efficient technique for highly correlated users which are prevalent in 5G channels and are also common for systems with small numbers of transmit antennas. The beauty of DZF is that it is a variant of ZF that provides robustness against the joint scheduling of correlated users with very little additional complexity. Analysis of the two user case clearly demonstrates the idea of DZF and illustrates the potential for improved rates and fairness. Both these performance gains are verified by simulation.

Chapter 4

Hybrid Zero-Forcing for Correlated and Semi-Orthogonal MU-MIMO Channels

In chapter 3, we found that scheduling correlated users and using ZF degrades the system performance due to noise inflation during channel inversion. To reduce this noise inflation, DZF has been proposed which decorrelates the channels of two highly correlated users. The proposed DZF design is similar to CZF in that for far users and unpaired users, the BS performs ZF and the users employ the leading left singular vectors for receive processing [33]. Unlike CZF, the near users employ the second strongest left singular vector for receive processing. As near and far users have highly correlated channels and the first two singular vectors are orthogonal, DZF aims for decorrelation (orthogonalization) of the near and far user effective channels, hence reducing noise inflation [23].

In practical scenarios, it is preferable to serve as many users as possible to reduce user latency and improve user fairness [103]. For example, users that contain delay constrained traffic may not be able to be scheduled in the next time slot [33] [39]. Moreover, users are often classified according to their spatial separability and this is done by quantifying the correlation among the users. Only the users whose mutual correlation is sufficiently low are allowed to use CZF [33] [39]. In contrast, the users

whose channel correlations are high employ DZF. DZF provides robustness against the joint scheduling of correlated users. Thus, in this chapter, on the basis of the advantages of both CZF and DZF, we propose a hybrid ZF (HZF) algorithm that selects both highly correlated and semi-orthogonal users. HZF selects highly correlated users on the basis of a correlation threshold value (Ψ_0) and also makes sure that the HZF rate is higher than the corresponding CZF rate to achieve higher sum-rate than CZF.

Note that in sections 4.1, 4.1.1, and 4.1.2, for completeness we summarize key information from chapters 2 and 3.

4.1 System Model

Consider a MU-MIMO system in which the BS is equipped with M_T transmitting antennas. The BS supports N UEs each with M_R receiving antennas. We assume spatially correlated Rician fading channels with unequal statistics for each user. Hence, the channel for user i can be written as

$$\mathbf{H}_i = \sqrt{\beta_i}\mathbf{G}_i, \tag{4.1}$$

where β_i is the link gain and $\mathbf{G}_i$ is the normalized channel denoted by

$$\mathbf{G}_i = \sqrt{\frac{K_i}{K_i+1}}\ \bar{\mathbf{H}}_i + \sqrt{\frac{1}{K_i+1}}\hat{\mathbf{H}}_i\mathbf{R}_i^{1/2}, \tag{4.2}$$

where $\bar{\mathbf{H}}_i$, $\hat{\mathbf{H}}_i$ and $\mathbf{R}_i$ are the specular (LOS), diffuse (scattered) components of the channel and the $M_T \times M_T$ transmit spatial correlation matrix, respectively. The ratio between the powers of the specular and scattered components is defined as the Rician (K) factor. Note that K_i and $\mathbf{R}_i$ are unique to each user. Assuming a uniform linear array (ULA) at both ends of the link, the LOS component is an $M_R \times M_T$ matrix

given by

$$\bar{\mathbf{H}}_i = \begin{bmatrix} 1 \\ e^{j\,2\pi d \sin\bar{\phi}_{iR}} \\ \vdots \\ e^{j\,2\pi(M_R-1)d\sin\bar{\phi}_{iR}} \end{bmatrix} \times \begin{bmatrix} 1 & e^{j\,2\pi d\sin\bar{\phi}_{iT}} & \cdots & e^{j\,2\pi(M_T-1)d\sin\bar{\phi}_{iT}} \end{bmatrix}, \tag{4.3}$$

where $\bar{\phi}_{iR}$ is the LoS angle-of-arrival (AoA) at the receiver, $\bar{\phi}_{iT}$ is the LoS angle-of-departure (AoD) at the transmitter, and d is the inter-element spacing in wavelengths. The SVD of the channel is denoted $\mathbf{H}_i = \mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^H$, where $\mathbf{U}_i$ and $\mathbf{V}_i$ are $M_R \times M_R$ and $M_T \times M_T$ unitary matrices, respectively, and $\mathbf{D}_i = \text{diag}(d_1^{(i)}, d_2^{(i)}, \cdots, d_m^{(i)})$ contains the $m = \min(M_R, M_T)$ ordered singular values of $\mathbf{H}_i$ such that $d_1^{(i)} \geq d_2^{(i)} \geq \cdots \geq d_m^{(i)}$.

We model the distribution of users with a uniform density over the coverage area of the cell as in [104]. The link gains are $\beta_i = A\xi_i(r_0/r_i)^{-\gamma}$, where A is the unitless constant for the geometric attenuation at a reference distance of r_0, r_i is the distance between the BS and the i^{th} user, γ is the attenuation exponent and $\xi_i = 10^{(X_i/10)}$, where X_i depends upon the shadow fading standard deviation (σ_{sf}^2), i.e., $X_i \sim \mathcal{N}(0, \sigma_{sf}^2)$.

4.1.1 Conventional precoder design

As we discussed in Sec. 3.2.1, a conventional approach to delivering one stream per user is to use the dominant left singular vector of the channel at each user and plain ZF at the BS. We refer to this technique as CZF. Denoting the first column of $\mathbf{U}_i$ by $\mathbf{u}_i$, the effective user channels are then $\mathbf{h}_i = \mathbf{u}_i^H\mathbf{H}_i$, $i = 1, \cdots N$ and the global channel from the BS to the N users is

$$\mathbf{H} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_N^T]^T. \tag{4.4}$$

The transmitter performs ZF based on $\mathbf{H}$, assuming $M_T \geq N$. The un-normalized precoding matrix, $\mathbf{W}_0$, is given by

$$\mathbf{W}_0 = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1} = [\mathbf{w}_{01}, \mathbf{w}_{02}, \cdots \mathbf{w}_{0N}], \tag{4.5}$$

so that the normalized precoder is $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N]$, where $\mathbf{w}_i = \mathbf{w}_{0i}/||\mathbf{w}_{0i}||$.

4.1.2 DZF precoder design

As we discussed in Sec. 3.2.2, the DZF scheme groups the $N = 2N_1 + N_2$ users into N_1 pairs (each containing a highly correlated near (s) and far (w) user) and N_2 singletons (users that can't be grouped into a highly correlated pair). Accordingly we relabel the channel matrices so they are ordered as $\mathbf{H}_{is}, \cdots, \mathbf{H}_{N_1 s}, \mathbf{H}_{1w}, \cdots, \mathbf{H}_{N_1 w}$, $\mathbf{H}_{N_1+1}, \cdots, \mathbf{H}_N$ and each user pair, i, contains the highly correlated channels $(\mathbf{H}_{is}, \mathbf{H}_{iw})$. DZF is similar to CZF in that the BS performs traditional ZF and the singletons and far users employ the leading left singular vectors for processing. In contrast, the near users employ the second strongest left singular vector which decorrelates their channel with the paired far user at the expense of a loss in power. Hence, the users perform receive processing to create effective channels as follows:

$$\mathbf{h}_{is} = \mathbf{u}_{is}^H \mathbf{H}_{is}, \;\; \mathbf{u}_{is} = \text{col. 2 of } \mathbf{U}_{is}, \;\; \mathbf{H}_{is} = \mathbf{U}_{is}\mathbf{D}_{is}\mathbf{V}_{is}^H, \tag{4.6}$$

$$\mathbf{h}_{iw} = \mathbf{u}_{iw}^H \mathbf{H}_{iw}, \;\; \mathbf{u}_{iw} = \text{col. 1 of } \mathbf{U}_{iw}, \;\; \mathbf{H}_{iw} = \mathbf{U}_{iw}\mathbf{D}_{iw}\mathbf{V}_{iw}^H, \tag{4.7}$$

for $i = 1, \cdots, N_1$ and

$$\mathbf{h}_i = \mathbf{u}_i^H \mathbf{H}_i, \;\; \mathbf{u}_i = \text{col. 1 of } \mathbf{U}_i, \;\; \mathbf{H}_i = \mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^H, \tag{4.8}$$

for $i = N_1 + 1, \cdots N$. Now, the global equivalent channel matrix, $\mathbf{Q}$, considering all users is given by

$$\mathbf{Q} = [\mathbf{h}_{is}^T, \cdots, \mathbf{h}_{N_1 s}^T, \mathbf{h}_{1w}^T, \cdots, \mathbf{h}_{N_1 f}^T, \mathbf{h}_{N_1+1}^T, \cdots, \mathbf{h}_N^T]^T. \tag{4.9}$$

The un-normalized ZF precoder is $\widetilde{\mathbf{W}}_0 = \mathbf{Q}^H(\mathbf{Q}\mathbf{Q}^H)^{-1} = [\widetilde{\mathbf{w}}_{01}, \widetilde{\mathbf{w}}_{02}, \cdots, \widetilde{\mathbf{w}}_{0N}]$, and the normalized precoder, $\widetilde{\mathbf{W}} = [\widetilde{\mathbf{w}}_1, \widetilde{\mathbf{w}}_2, \cdots \widetilde{\mathbf{w}}_N]$, is calculated similarly, where $\widetilde{\mathbf{w}}_i = \widetilde{\mathbf{w}}_{0i}/||\widetilde{\mathbf{w}}_{0i}||$.

4.1.3 SNR calculations for the two user case

In chapter 3, we found that the DZF scheme is better than CZF when highly correlated users are present. However, the channel powers are also critical in determining the performance of the systems. Thus, to determine the effect of both channel powers and correlation we derive the SNRs of the two user case.

Considering the two user case, $\mathbf{H}_{eff} = [\mathbf{h}_1^T, \mathbf{h}_2^T]^T$, where $\mathbf{h}_1$ and $\mathbf{h}_2$ are the effective channels of the users either from CZF or DZF. Employing ZF, the SNR of both users can be obtained from (3.18) in chapter 3 as

$$\mathrm{SNR}_i = \rho \left(\left[\left(\mathbf{H}_{eff} \mathbf{H}_{eff}^H \right)^{-1} \right]_{ii} \right)^{-1}, \tag{4.10}$$

where $(\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}$ can be calculated as

$$\mathbf{H}_{eff}\mathbf{H}_{eff}^H = \begin{bmatrix} \mathbf{h}_1\mathbf{h}_1^H & \mathbf{h}_1\mathbf{h}_2^H \\ \mathbf{h}_2\mathbf{h}_1^H & \mathbf{h}_2\mathbf{h}_2^H \end{bmatrix}, \tag{4.11}$$

so that

$$(\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1} = \frac{\begin{bmatrix} \mathbf{h}_2\mathbf{h}_2^H & -\mathbf{h}_1\mathbf{h}_2^H \\ -\mathbf{h}_2\mathbf{h}_1^H & \mathbf{h}_1\mathbf{h}_1^H \end{bmatrix}}{\mathbf{h}_1\mathbf{h}_1^H\mathbf{h}_2\mathbf{h}_2^H - \mathbf{h}_1\mathbf{h}_2^H\mathbf{h}_2\mathbf{h}_1^H}. \tag{4.12}$$

Using (4.10), the SNRs can be computed as

$$\begin{aligned}
\mathrm{SNR}_1 &= \frac{\rho}{\left(\frac{\mathbf{h}_2\mathbf{h}_2^H}{\mathbf{h}_1\mathbf{h}_1^H\mathbf{h}_2\mathbf{h}_2^H - \mathbf{h}_1\mathbf{h}_2^H\mathbf{h}_2\mathbf{h}_1^H}\right)} \\
&= \rho\left(\mathbf{h}_1\mathbf{h}_1^H - \frac{\mathbf{h}_1\mathbf{h}_2^H\mathbf{h}_2\mathbf{h}_1^H}{\mathbf{h}_2\mathbf{h}_2^H}\right) \\
&= \rho\mathbf{h}_1\mathbf{h}_1^H\left(1 - \frac{|\mathbf{h}_1\mathbf{h}_2^H|^2}{(\mathbf{h}_1\mathbf{h}_1^H)(\mathbf{h}_2\mathbf{h}_2^H)}\right) \\
&= \rho\varkappa_1(1 - |\Psi_{12}|^2),
\end{aligned} \tag{4.13}$$

where $\varkappa_1 = \mathbf{h}_1\mathbf{h}_1^H$ is the channel power of user 1 and $\Psi_{12} = \frac{|\mathbf{h}_1\mathbf{h}_2^H|}{\sqrt{\left((\mathbf{h}_1\mathbf{h}_1^H)(\mathbf{h}_2\mathbf{h}_2^H)\right)}}$ is the correlation between the channels of users 1 and 2. Similarly, SNR_2 can be calculated as

$$\mathrm{SNR}_2 = \rho\varkappa_2(1 - |\Psi_{12}|^2), \tag{4.14}$$

where $\varkappa_2 = \mathbf{h}_2\mathbf{h}_2^H$ is the channel power of user 2. From the above SNR computations we can conclude that the SNRs are directly proportional to channel power, however SNR decreases with an increase in correlation. Using the DZF technique, $\varkappa_1$ is reduced but also $|\Psi_{12}|^2$ is reduced if the correlation was originally higher. From this we can conclude that we should take both power and correlation into account when deciding on DZF versus CZF.

For simplicity, let us denote the correlation of two users in CZF as Ψ_C, which can be written as

$$\Psi_C = |\mathbf{h}_{1C}\mathbf{h}_2^H| / \sqrt{\left((\mathbf{h}_{1C}\mathbf{h}_{1C}^H)(\mathbf{h}_2\mathbf{h}_2^H)\right)}, \tag{4.15}$$

where the effective channel, $\mathbf{h}_{1C}$, employs the first eigen vector. Similarly, for DZF, the correlation of two users can be written as

$$\Psi_D = |\mathbf{h}_{1D}\mathbf{h}_2^H| / \sqrt{\left((\mathbf{h}_{1D}\mathbf{h}_{1D}^H)(\mathbf{h}_2\mathbf{h}_2^H)\right)}, \tag{4.16}$$

where the stronger users effective channel, $\mathbf{h}_{1D}$, employs the second eigen vector. In

both cases, the second or weak user's effective channel, $\mathbf{h}_2$, employs the first eigen vector. Therefore, for CZF, the SNRs of both users can be written as

$$\mathrm{SNR}_{1C} = \rho\varkappa_{1C}(1 - |\Psi_C|^2), \text{ and } \mathrm{SNR}_{2C} = \rho\varkappa_2(1 - |\Psi_C|^2). \tag{4.17}$$

Similarly, for DZF, the SNRs of both users can be written as

$$\mathrm{SNR}_{1D} = \rho\varkappa_{1D}(1 - |\Psi_D|^2), \text{ and } \mathrm{SNR}_{2D} = \rho\varkappa_2(1 - |\Psi_D|^2). \tag{4.18}$$

The sum-rates for CZF and DZF can then be computed as

$$\mathrm{R}_C = \log_2(1 + \mathrm{SNR}_{1C}) + \log_2(1 + \mathrm{SNR}_{2C}), \tag{4.19}$$

and

$$\mathrm{R}_D = \log_2(1 + \mathrm{SNR}_{1D}) + \log_2(1 + \mathrm{SNR}_{2D}), \tag{4.20}$$

respectively. These derivations and observations are used to develop the HZF schemes described in section 4.1.4.

4.1.4 Proposed hybrid zero-forcing (HZF) algorithms

In chapter 3 we discussed CZF and DZF and their potential to achieve better performance in different user correlation environments. However, we have not discussed any methods to schedule the users so that both CZF and DZF can be best utilized to achieve higher system performance. Therefore, in this section we propose two algorithms that integrate both CZF and DZF schemes to form: HZF with correlation only (HZF-C) and HZF with correlation and channel power (HZF-CP) which are described below. The main idea of both algorithms is to provide fairness while scheduling, i.e., to schedule both semi-orthogonal and highly correlated users as required, and minimize the user selection complexity. Further, both algorithms achieve higher performance. For comparison purposes, we also discuss HZF with an exhaustive search method (HZF-E) that gives an upper bound on the system performance.

HZF with correlation only (HZF-C) algorithm

As we have already discussed in Sec. 2.5 that the BS can classify the users according to their spatial separability which can be quantified by the correlation between the spatial signatures of the users [98]. We can measure the instantaneous channel similarity of the users by correlating the channel vectors. Moreover, we know that if the channels of the selected users have a correlation below a certain threshold value, CZF is used [33]. Furthermore, when the selected users are highly correlated then DZF is superior to CZF. Thus using the correlation factor and leverage the potential of both CZF and HZF scheme, we propose the HZF-C algorithm as given below.

1. L users are selected for transmission at a given time on the basis of CSI feedback to the transmitter. Hence, $\mathbf{H}_1, \mathbf{H}_2 \cdots, \mathbf{H}_L$ are known at the transmitter. Initialize $r = 1$ and go to step 2.

2. For i $= 1, \ldots, L$, compute the SVD $\mathbf{H}_i = \mathbf{U}_i \mathbf{D}_i \mathbf{V}_i^H$, and form the effective channels

$$\mathbf{u}_i^H \mathbf{H}_i = \mathbf{D}_{11}^{(i)} \mathbf{v}_i^H. \tag{4.21}$$

   The effective channel is proportional to $\mathbf{v}_i$, where $\mathbf{v}_i$ = col. 1 of $\mathbf{V}_i$. Thus, compute $\mathbf{v}_1, \cdots, \mathbf{v}_L$ and the resulting matrix of cross-products

$$\mathbf{M} = \begin{bmatrix} \mathbf{v}_1^H \\ \vdots \\ \mathbf{v}_L^H \end{bmatrix} \times \begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_L \end{bmatrix}. \tag{4.22}$$

   For example, let $L = 4$ then after computation, the upper triangular part of $\mathbf{M}$ can be written as

$$\mathbf{M} = \begin{bmatrix} 1 & \mathbf{v}_1^H \mathbf{v}_2 & \mathbf{v}_1^H \mathbf{v}_3 & \mathbf{v}_1^H \mathbf{v}_4 \\ . & 1 & \mathbf{v}_2^H \mathbf{v}_3 & \mathbf{v}_2^H \mathbf{v}_4 \\ . & . & 1 & \mathbf{v}_3^H \mathbf{v}_4 \\ . & . & . & 1 \end{bmatrix}. \tag{4.23}$$

3. The cross product and upper triangular part of M can be rewritten as

$$\mathbf{M} = \begin{bmatrix} 1 & C_{12} & C_{13} & C_{14} \\ . & 1 & C_{23} & C_{24} \\ . & . & 1 & C_{34} \\ . & . & . & 1 \end{bmatrix}. \tag{4.24}$$

4. Order the absolute values of the C_{ij} terms in $C = \Big\{|C_{12}|, |C_{13}|, \cdots, |C_{(L-1)L}|\Big\}$. Further, select the largest element of C, denoted $C_{i_r j_r}$.

5. If $C_{i_r j_r} > \Psi_0$, use DZF for users i_r, j_r.
   Else use CZF for users i_r, j_r.

6. Remove all terms in $\mathbf{M}$ corresponding to users i_r and j_r and repeat the process for r = r + 1.

7. Stop when no two users have a correlation which is greater than Ψ_0.

HZF using both correlation and channel power(HZF-CP) algorithm

The HZF-C algorithm described above is based on the correlation among the users. However, as we have observed in Sec. 4.1.3 channel power is also important when deciding on DZF versus CZF. Thus, we have added a step in HZF-C as below to construct the HZF-CP algorithm which may achieve higher performance than HZF-C.

- Repeat steps 1 - 4 of HZF-C algorithm.
- Step 5
  If $C_{i_r j_r} > \Psi_0$ and $R_D > R_C$ for users i_r and j_r
  - Use DZF for users i_r, j_r.
  Else use CZF for users i_r, j_r.
- Repeat steps 6 and 7 of HZF-C algorithm.

HZF exhaustive search (HZF-E) algorithm

The HZF-E search algorithm involves choosing all possible ways to allocate DZF and CZF to user pairs and checking for the best combination to maximize the sum-rate of the system. For example, for the two user case it selects the best combination of both users which either employs the dominant eigen vectors at both users as in CZF or employs the first eigen vector at the weak user and the second eigen vector at the strong user as in HZF or vice versa. Hence, there are three combinations and the method with the highest rate is chosen. The highest rate gives an upper bound on the sum-rate of the system. Similarly, for 3 users there are 4 combinations and for 4 users there are 11 combinations as given in Appendix A.1. Therefore, with an increase in the number of users, the search complexity of the system increases. Due to this high level of computational complexity in the exhaustive search method, we only consider $N \in \{2, 3, 4\}$ in the numerical results.

4.2 Numerical Results

We first compare the HZF-C method with the CZF and DZF techniques presented in chapter 3 to confirm that the HZF-C method can achieve performance gains. Similar to the system considered in chapter 3, we consider $M_T = 2$ antennas at the BS and $M_R = 2$ antennas at all users unless otherwise specified. To construct the channels of highly correlated users, we first generate an uncorrelated Rayleigh fading channel matrix for the near user, $\mathbf{H}_s$, with elements which are $\mathcal{CN}(0, \beta_s = 4)$. We then generate a correlated channel,

$$\mathbf{H}_w = \sqrt{\beta_w/\beta_s}(\Psi\mathbf{H}_s + \sqrt{1-\Psi^2}\mathbf{E}_s), \tag{4.25}$$

with elements which are $\mathcal{CN}(0, \beta_w = 1)$, where Ψ is the correlation between the near and far user and the elements of $\mathbf{E}_s$ are $\mathcal{CN}(0, \beta_s)$. This allows us to control correlation through Ψ. The remaining unpaired users have independent Rayleigh channel matrices with elements $\mathcal{CN}(0, \beta_i = 1)$.

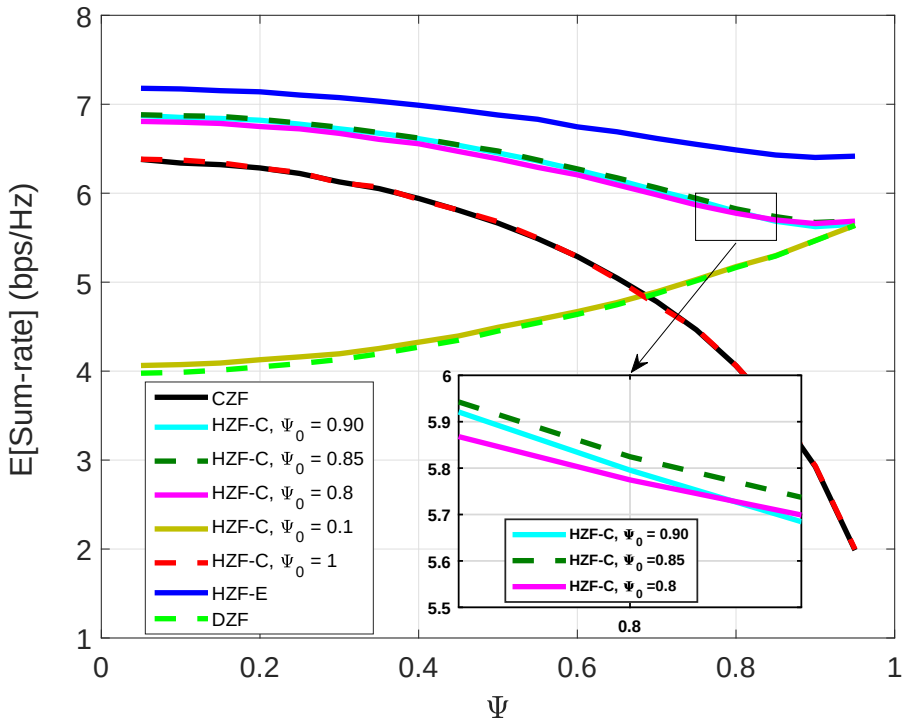


Figure 4-1: Sum-rate vs. correlation comparison for all schemes. $M_T = 2$, $N = 2$, $M_R = 2$, $\beta_s = 4$ and $\beta_w = 1$.

In Fig. 4-1, we can observe that HZF-C has a higher performance than both CZF and DZF schemes for all values of Ψ and a range of correlation threshold values (Ψ_0). This is due to the HZF design process, where the user similarity is checked using step 2 in the HZF-C algorithm and CZF or DZF is selected accordingly. We can observe that the proposed HZF-C method is identical to CZF when $\Psi_0 = 1$. This is because we never see perfectly correlated users in practice. Hence, the correlation never reaches $\Psi_0 = 1$ and CZF is employed for all users. In contrast, when $\Psi_0 = 0.1$ the HZF-C curve is close to DZF. Here, HZF-C mostly employs DZF as the correlation between users will be above the low correlation threshold value most of the time. In Fig. 4-1, we show that the HZF-C method is insensitive to the correlation threshold values ranging from $\Psi_0 = 0.80$ to $\Psi_0 = 0.90$. This conclusion is validated for a range of channels, powers and system sizes in Figs. 4-2, 4-3 and 4-5 to 4-11. Therefore, we can confirm that HZF is a robust technique that achieves higher sum-rate compared to CZF or DZF techniques.

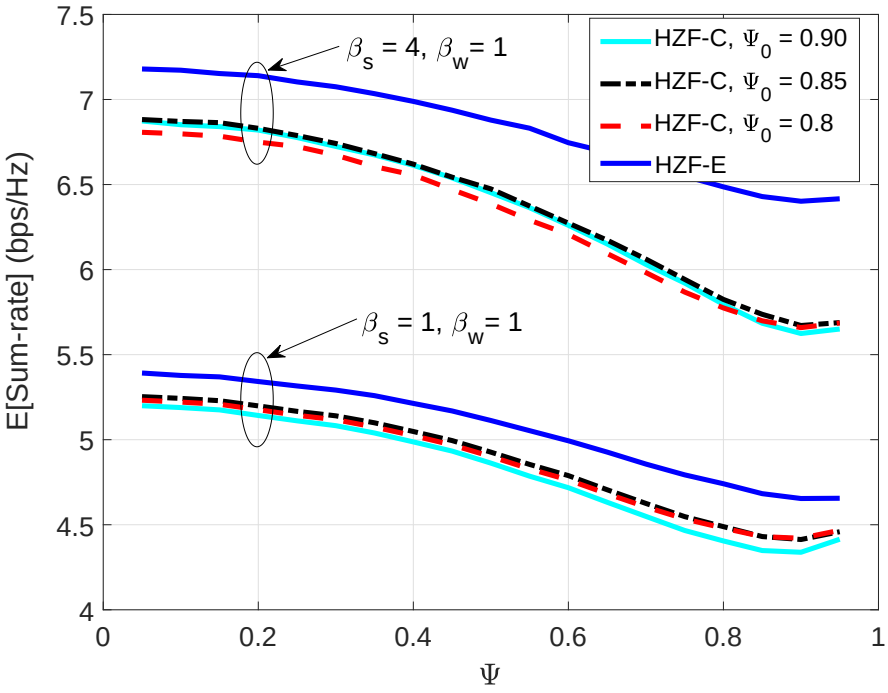


Figure 4-2: Sum-rate vs. correlation for HZF schemes. $M_T = 2$, $M_R = 2$, and $N = 2$.

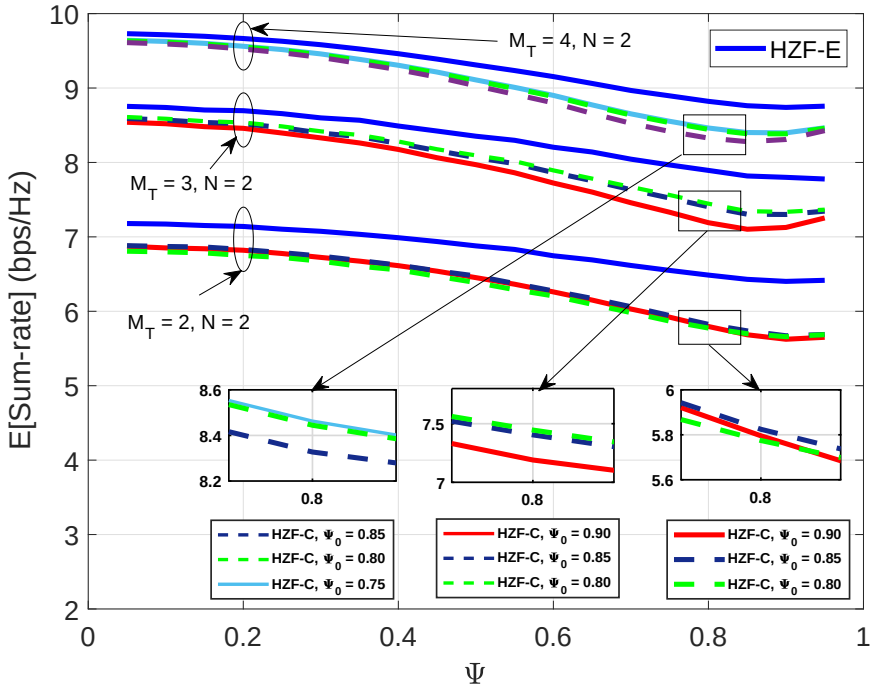


Figure 4-3: Sum-rate vs. correlation comparison for HZF schemes. $\beta_s = 4$ and $\beta_w = 1$.

In Figs. 4-2 and 4-3, as in chapter 3, we consider the performance of HZF for different system parameter values, such as the number of transmitting antennas (M_T) and user channel powers. Similar to the observations made in Figs. 3-4 and 3-5, we can observe that the sum-rate increases when the strong user's channel gain is increased from $\beta_s = 1$ to 4 and also with an increase in the number of transmitting antennas[1]. In Fig. 4-2, though we increase the strong user's channel gain from $\beta_s = 1$ to 4, the highest rate achieving threshold value remains constant which is $\Psi_0 = 0.85$. On the other hand, considering a fixed number of users $N = 2$ and increasing the number of transmitting antennas we observe that slightly higher rates can be achieved with a decrease in correlation threshold values, i.e., for $M_T = 2$, 3 and 4 we can observe that $\Psi_0 = 0.85$, 0.80 and 0.75 achieve slightly higher rates compared to other values, respectively.

In all cases, HZF-E has the highest performance compared to the other schemes. The interesting thing about the gains of the exhaustive search method is that they result from favoring the strong user. Simulations show that when HZF-E has a higher sum-rate than HZF it is almost always because the strong user was served with the leading eigen vector and the weak user employed the second eigen vector. This trend makes sense as it is common for the maximum sum-rate approach to be "greedy" and give more resources to the strong users. We denote this approach reverse DZF (R-DZF) as it reverses the eigen vector ordering used by DZF. In this thesis, R-DZF is not pursued further as it results in a considerable lack of fairness. For example, in Fig. 4-4 DZF and R-DZF are compared. While DZF gives good fairness and similar rates to the two users, R-DZF creates a very large performance gap.

Since system performance depends upon the channel environments, in order to fully investigate the effectiveness of the HZF methods we consider more realistic channel models, based on correlated Rician fast fading and log-normal shadowing with pathloss as given in Sec. 4.1. User distances are randomly generated, corresponding to a uniform spatial distribution, with a maximum distance of $r = 100$m and a reference

[1]For clarity in the figures, we omitted the CZF and DZF curves as the trend of the curves are already shown in Fig. 4-1 and HZF is always better than CZF and DZF methods.

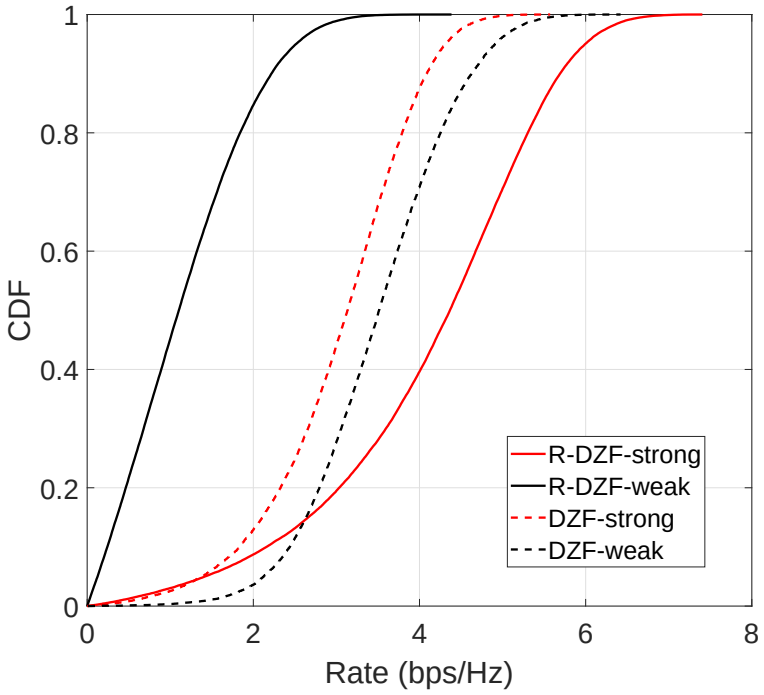
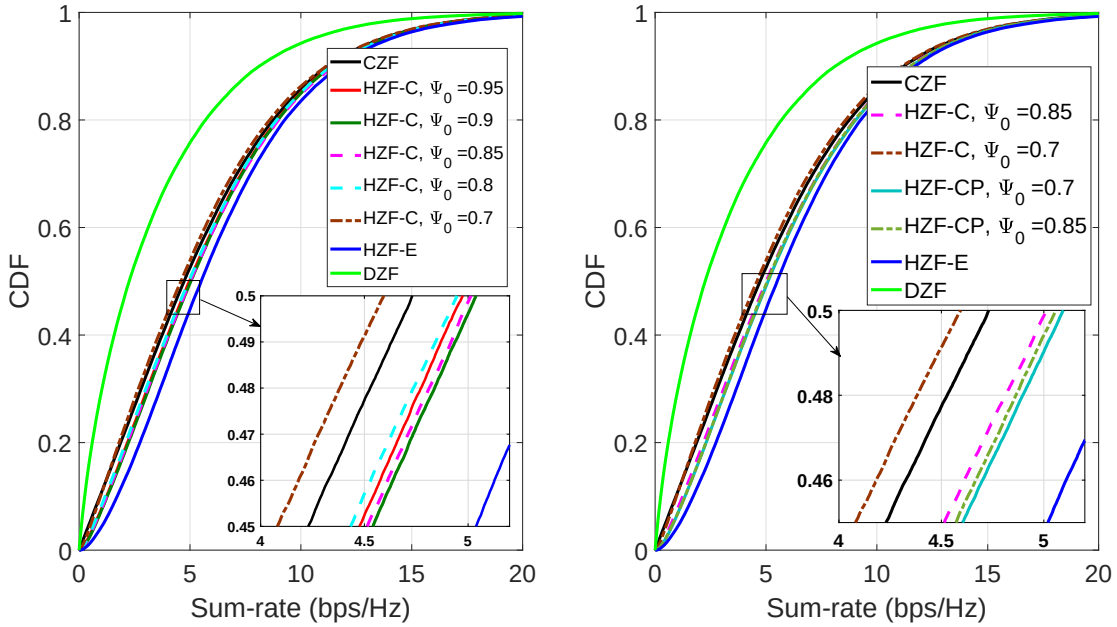


Figure 4-4: Rate CDF comparison of 2 users with DZF and R-DZF. $\rho = 5$ dB, $M_T = M_R = 2$, $N = 2$, $\beta_s = 4$ and $\beta_w = 1$.

distance $r_0 = 1$m. The other pathloss parameters are $\sigma_{sf} = 8$, and $\gamma = 3.5$, while the parameter A is chosen such that the tenth percentile of the instantaneous SNR(dB) of ZFBF with four users selected by SUS [33] is 0 dB, when $M_T = 4$ and $\rho = 15$ dB. Further, we employ the one ring model for the spatial correlation matrices defined in Sec. 2.1.3 given by (2.14). We set $d = 0.25$ and assume both fixed and user-specific ϕ values.

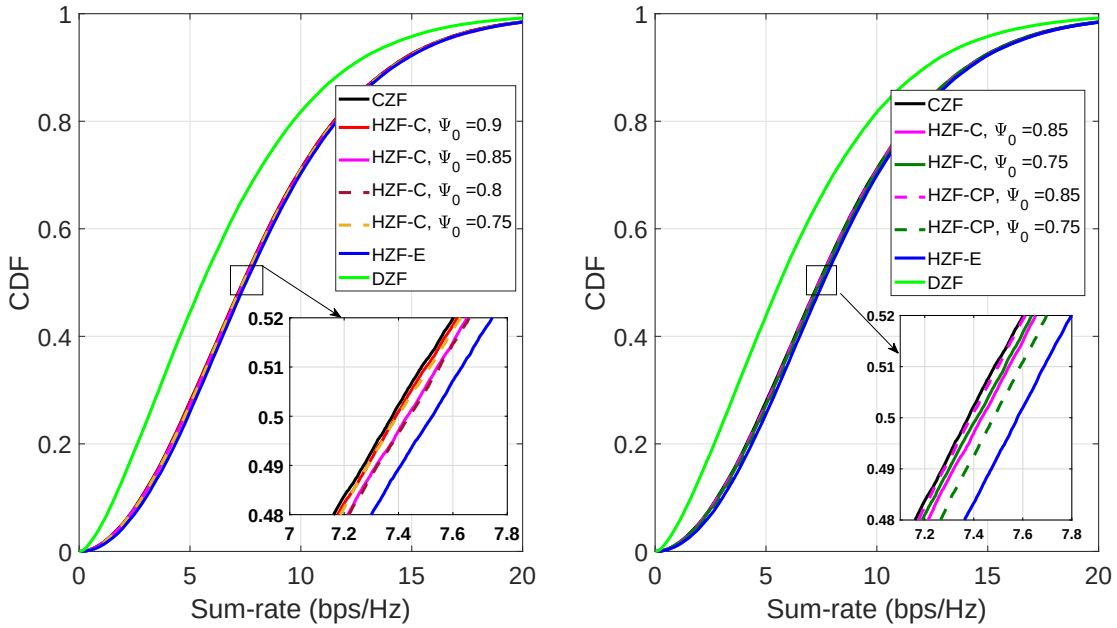
In Figs. 4-5 to 4-11 we simulate five methods: CZF, DZF, HZF-C, HZF-CP, and HZF-E in different channel environments. For each environment, we show the impact of the number of transmitting antennas and the correlation threshold values. In general, for both $M_T = 2$ and $M_T = 4$, we can observe that HZF-C has a performance gain over both CZF and DZF. Further, HZF-CP has a slightly higher performance than HZF-C. This is because HZF-CP benefits from both a power and correlation based cluster selection procedure using step 5 of the HZF-CP algorithm, while HZF-C only checks the correlation among the users and doesn't exploit the channel power.

We have already discussed that HZF is a robust technique as the system perfor-



(a) HZF-C (b) HZF-C vs HZF-CP

Figure 4-5: Sum rate CDFs of I.I.D. Rayleigh channel. $M_T = 2$, $M_R = 2$, $N = 2$, and $\rho = 5$ dB.

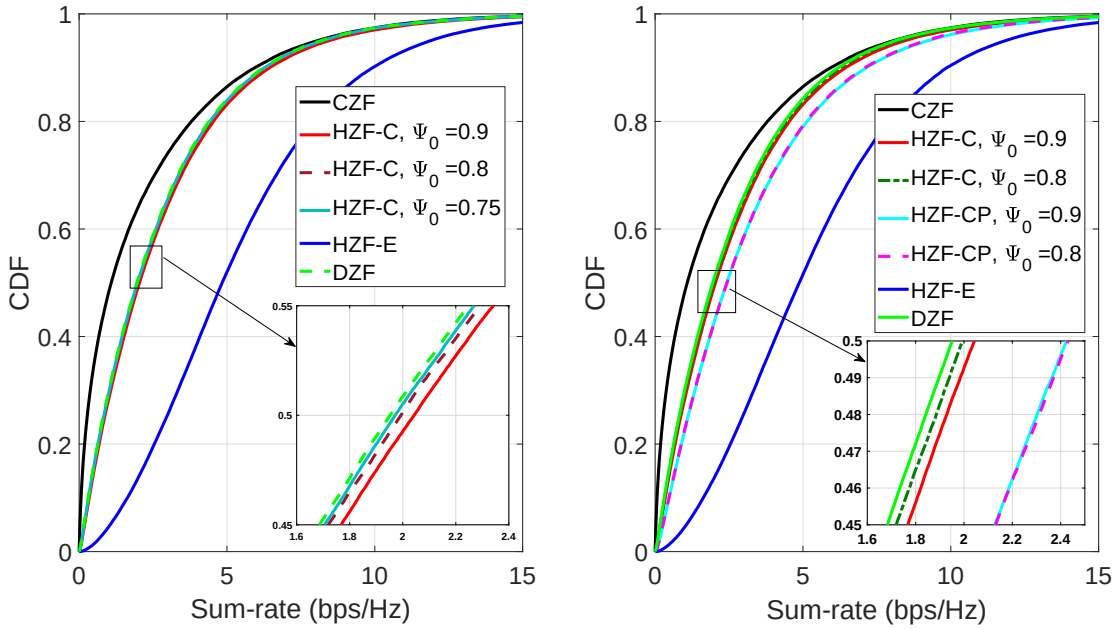


(a) HZF-C (b) HZF-C vs HZF-CP

Figure 4-6: Sum rate CDFs of I.I.D. Rayleigh channel. $M_T = 4$, $M_R = 2$, $N = 2$, and $\rho = 5$ dB.

mance is insensitive to Ψ_0. This is now validated for a range of channels and system sizes. In Figs. 4-5 to 4-11, we can observe that the gains of HZF-C are similar for a range of Ψ_0 values, i.e., $\Psi_0 = 0.80$ to $\Psi_0 = 0.90$. However, there are some small variations in the correlation threshold values in different environments that achieve slightly higher rates compared to other threshold values. In Fig. 4-5a, for $M_T = 2$ in I.I.D. Rayleigh channels we can observe that $\Psi_0 = 0.9$ achieves a slightly higher rate compared to other values. On the other hand in Fig. 4-6a for $M_T = 4$, we see that $\Psi_0 = 0.8$ achieves a slightly higher rate. In all figures, we observe a similar trend, namely that increases in correlation among users, slightly increases the correlation threshold point that achieves the highest rate.

Moreover, in Fig. 4-5b HZF-C has a higher performance when $\Psi_0 = 0.9$, however for HZF-CP it is $\Psi_0 = 0.7$. Hence, HZF-CP has a performance gain over HZF-C for comparatively lower correlation threshold values. This is reasonable as the power comparison can be more useful at lower correlations. We know that the performance

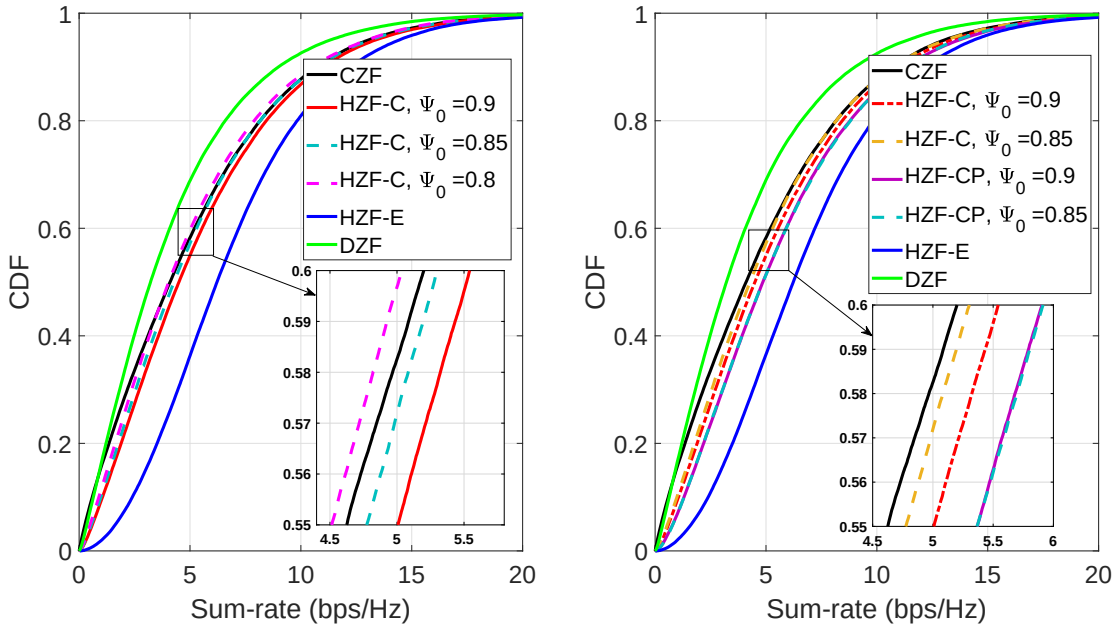


(a) HZF-C (b) HZF-C vs HZF-CP

Figure 4-7: Sum rate CDFs of Corr. Rayleigh channel. $M_T = 2$, $M_R = 2$, $N = 2$, and $\rho = 5$ dB.

of the systems is environment dependent and in our simulations this is shown with

three key parameters: a) the number of transmit antennas, b) the LoS factor (K) and c) the spatial correlation among users. These three parameters also play a vital role in the performance gap between CZF and DZF techniques and further this has an impact on the gains of the HZF techniques. In Figs. 4-5, 4-6, 4-8, and 4-10, we can observe that DZF has the lowest rate compared to the other techniques. On the other hand, in Figs. 4-7, 4-9, and 4-11 CZF has the lowest rate compared to the other techniques. Hence, it is valuable to have a hybrid technique which selects the best option in any environment.

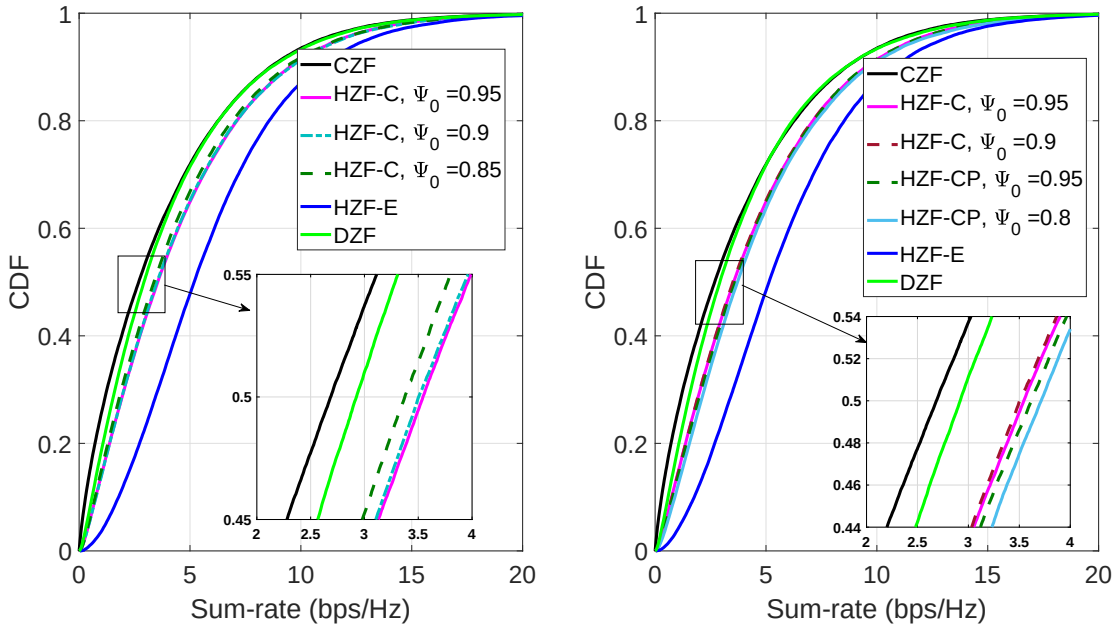


(a) HZF-C (b) HZF-C vs HZF-CP

Figure 4-8: Sum rate CDFs of Corr. Rayleigh channel. $M_T = 4$, $M_R = 2$, $N = 2$, and $\rho = 5$ dB.

Firstly, we observe in Fig. 4-7 that for $M_T = 2$ in a correlated Rayleigh channel, DZF has a higher rate than CZF. However, in Fig. 4-8, when the number of transmit antennas are increased to $M_T = 4$, CZF has a higher rate than DZF. This is intuitive since increasing the number of antennas decreases the correlation among the users which favors CZF performance. Secondly, we can observe a similar trend while comparing Figs. 4-9 and 4-10. However, the performance gap between CZF and DZF decreases in I.I.D. Rician channels compared to I.I.D. Rayleigh and Correlated

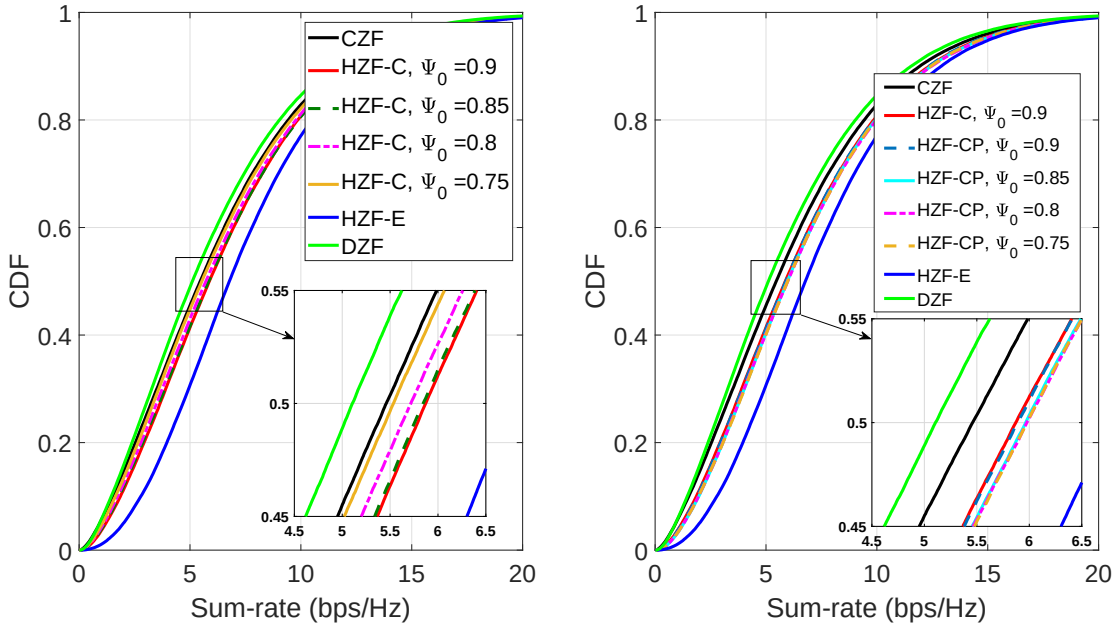
Rayleigh channels. This is because the K factor favors the DZF technique by increasing the correlation among the users. Finally, with spatial correlation in a correlated Rayleigh channel in Fig. 4-7 we can observe that DZF has a higher rate than CZF in contrast to Fig. 4-5. Moreover, the performance gap between the CZF and DZF decreases with spatial correlation, and this can be observed by comparing Figs. 4-6 with 4-8. As expected, with spatial correlation there are more highly correlated users, which increases the DZF performance, which further helps to reduce the performance gap between CZF and DZF.



(a) HZF-C (b) HZF-C vs. HZF-CP

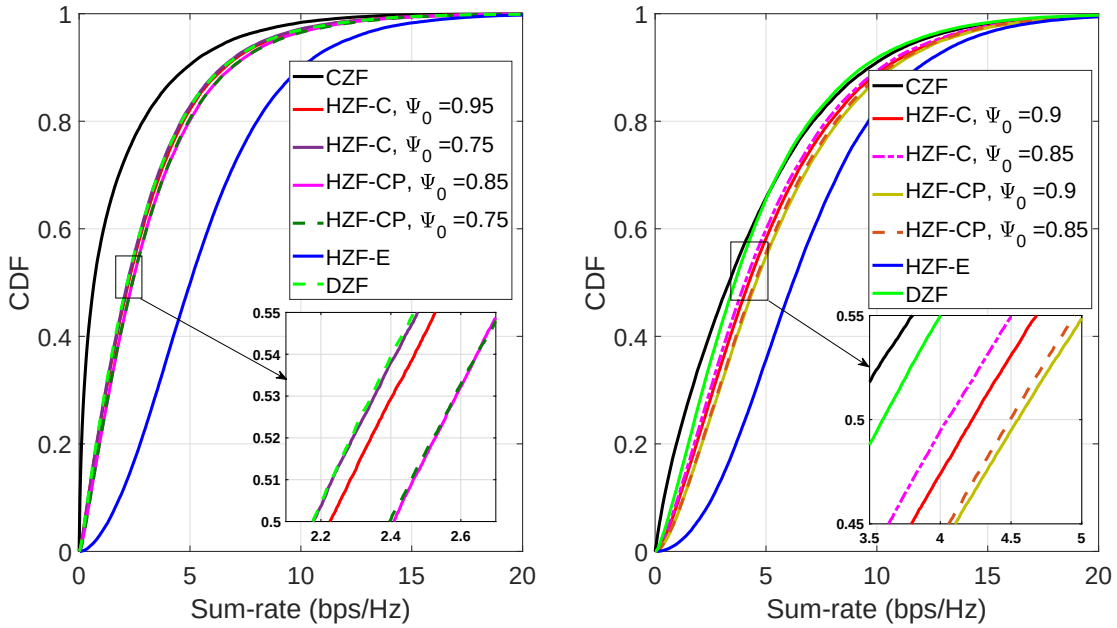
Figure 4-9: Sum rate CDFs of I.I.D. Rician channel. $M_T = 2$, $M_R = 2$, $N = 2$, and $\rho = 5$ dB.

When we compare the performance gap between HZF-C or HZF-CP and HZF-E methods the highest performance gap can be seen in correlated Rayleigh and then in correlated Rician channels with $M_T = 2$ antennas. In contrast, the lowest performance gaps can be seen in the I.I.D. Rayleigh channel with $M_T = 4$ antennas and then with $M_T = 2$ antennas. In Figs. 4-7 and 4-11a for $M_T = 2$ in correlated Rayleigh and correlated Rician channels, as expected there are more highly correlated users as observed during simulation. Further, when R-DZF is used in the presence of highly



(a) HZF-C (b) HZF-C vs. HZF-CP

Figure 4-10: Sum rate CDFs of I.I.D. Rician channel. $M_T = 4$, $M_R = 2$, $N = 2$, and $\rho = 5$ dB.



(a) HZF-C vs HZF-CP, $M_T = 2$ (b) HZF-C vs. HZF-CP, $M_T = 4$

Figure 4-11: Sum rate CDFs of Corr. Rician channel. $M_R = 2$, $N = 2$, and $\rho = 5$ dB.

correlated users it creates a large performance gap. This is due to the differences in the eigen values of the correlated and uncorrelated users, i.e., the gap in eigen values of correlated users is larger than that for uncorrelated users. Further, this gap decreases with an increase in the number of uncorrelated users, i.e., I.I.D. Rayleigh channels are more uncorrelated with more transmit antennas.

4.3 Conclusion

In this chapter, we integrated DZF and CZF schemes to leverage the potential of both schemes and referred to this combination as the HZF scheme. HZF is a simple yet efficient algorithm that provides fairness while scheduling the users and achieves performance gains. We have proposed two algorithms; HZF with correlation based selection only (HZF-C) and HZF with correlation and channel power based selection (HZF-CP). The HZF-C achieves higher performance than both CZF and DZF schemes for all values of Ψ. Also the HZF-C method is robust to the correlation threshold values. HZF-CP is slightly better than HZF-C for most of the channel environments, however it has higher system complexity. Simulations in I.I.D./ correlated Rayleigh and Rician channel environments clearly demonstrate the idea of the HZF scheme and illustrates the potential for improved rates compared to CZF and DZF schemes.

Chapter 5

NOMA with Guaranteed Weak User QoS: Design and Analysis

5.1 Introduction

In chapters 3 and 4, we discussed DZF and HZF signal processing schemes. The former technique simply decorrelates the channels of two highly correlated users exploiting the advantage of partial orthogonality of eigenvectors of highly correlated users. The latter approach is a robust technique while scheduling both semi-orthogonal and highly correlated users. However, with the emergence of stringent spectrally efficient system requirements for next generation cellular communication, a power domain multiplexing scheme has been proposed, referred to as NOMA.

NOMA is a power domain multiplexing technique, which employs superposition coding at the transmitter and SIC at one of the receivers [105] [106] [107]. Since several users' signals can be sent at the same time on the same frequency, more users can be accommodated in NOMA which increases the spectral efficiency of the system [26]. Furthermore, the concept of MU-MIMO systems can be incorporated in NOMA systems to enhance the system performance [108][109][110]. Multi-user ZFBF for a downlink MIMO-NOMA system in highly correlated user scenarios has been discussed in [40], [45] and [26]. In these papers, the users are grouped into a cluster based on their channel correlation and gain differences. Further, signals for these

highly correlated users are transmitted in a single beamforming vector. The work on NOMA-ZFBF systems in [40] and [45] showed that clustering highly correlated users is more likely to suppress the inter-cluster interference which in turn helps to increase the spectral efficiency. Similarly, [26] focused on the channel gains of paired users in a cluster and showed that the performance gain of NOMA is reduced when the users' channel gains in a cluster are similar. Moreover, [96] and [111] have studied clustering, power selection and precoding techniques. They achieve higher sum rates than [40]. [96] has considered the semiorthogonal user selection (SUS) algorithm from [33] and adopted its own user matching (UM) technique, referred to here as the SUS-UM algorithm. Thus, it was verified that with a combination of user clustering and power allocation techniques, a rate maximization scheme can be designed to achieve good performance for a large number of users.

The power allocation in a cluster is related to user fairness and quality of service (QoS) requirements [79][112]. In [79], resource allocation using proportional fairness was considered and [112] studied the power allocation coefficient choice using the concept of user fairness. In [113], the ergodic capacity maximization problem was studied under a total power constraint and a weak user minimum rate constraint. NOMA users' targeted data rates and allocated powers were found to be critical to their outage performance in [84]. [114] has implemented a predefined QoS for the weaker user's channel by choosing power allocation coefficients. Recently, [53] has proposed a dynamic power allocation scheme that accommodates flexible QoS requirements of clustered users. Therefore, power allocation and guaranteed QoS are the constraints to be solved during the design of NOMA systems. In addition, the spectral efficiency gains promised by NOMA-clusters are only achieved when the NOMA-cluster rate exceeds the TDMA rate [26] [46] [115].

Note that CZF, DZF and HZF schemes support each user by a single BF vector, in contrast NOMA-ZFBF involves designing a single BF vector to support two or more highly correlated users. When implementing the NOMA-ZFBF system performance depends upon the selection of the users in clusters and designing the ZFBF vector for each stream. Similar to the pairs referred to in chapters 3 and 4, NOMA-ZFBF also

groups two highly correlated users into pairs, where paired users are also referred to as near/strong users and far/weak users. Users not paired are referred to as singletons.

In this chapter, we focus on a downlink NOMA system, where a new power allocation strategy is proposed that ensures the weak user QoS (W-QoS). We refer to this as the W-QoS based NOMA (W-QoS NOMA) algorithm which is described in Sec. 5.2.5. Unlike other techniques, such as [40] [96], the proposed W-QoS NOMA technique provides both a minimum required QoS for the weak user and guarantees a NOMA rate higher than the TDMA rate. Moreover, the proposed W-QoS NOMA technique is very simple and easy to merge with established user selection and clustering algorithms. Thus, to implement the proposed W-QoS NOMA algorithm in a practical scenario, we choose two clustering algorithms, correlated user clustering (CUC) from [40] and SUS-UM from [96]. We modify the CUC algorithm [40] for implementation with W-QoS NOMA as described in Sec. 5.2.6. The results show that W-QoS NOMA merged with established user selection and clustering techniques has a performance gain over conventional techniques. Despite the fact that both SUS-UM and modified CUC use the same SUS based user selection approach, SUS-CUC has a marginal gain due to its clustering approach. Furthermore, we investigate the effect of the correlation factor on the cluster rate and observe that a high correlation among users in a cluster increases the cluster rate.

A closed form SNR/SINR analysis of NOMA systems with complex scheduling, power allocation, sum rate maximization and BF algorithms is intractable. Analytical progress was only possible for SIC detection and sum rate maximization by assuming perfectly correlated strong and weak user channels in a cluster, denoted a perfectly correlated user cluster. To the best of the author's knowledge, this is the only analysis of a NOMA-ZFBF system which includes the power allocation, guarantees the NOMA rate is higher than TDMA rate and gives a closed form upper bound for SNRs/SINRs and rate.

We have performed the analysis in two stages as described below:

(a) First, we analyze the SNR/SINR expressions for the strong and weak users in a cluster considering an arbitrary correlation between the users. This general

analysis leads to insights such as 1) the relationship between SNR/SINR and system parameters, such as transmit SNR, the correlation between users in a cluster, and the number of candidate users; 2) the importance of correlation between users in a cluster as the power penalty of the strong user and inter-cluster interference (ICI) / inter-user interference (IUI) in a cluster are directly proportional to the correlation factor; 3) unlike other techniques [26] [40] [114], it is interesting that W-QoS NOMA does not need a channel power differential between the strong and weak users in a cluster.

(b) Assuming perfect correlation, we derive the exact SNR/SINR distributions of singletons and strong and weak users in clusters which provides the closed form bounds. Moreover, on the basis of the derived SNR/SINR distribution results, we compute the expected sum rate of W-QoS NOMA in closed form in Sec. 5.4.3. This analytical solution for the sum rate gives an upper bound on the system performance. It is also useful in showing how the correlation factor affects the cluster rate and how close partially correlated channels can get to the perfectly correlated bound.

In section 5.2, we describe the NOMA-BF system model, W-QoS NOMA algorithm and SUS-CUC user selection and clustering algorithm. A performance analysis of the W-QoS NOMA algorithm with different selection and clustering algorithms is presented in section 5.3. Exact SNR/SINR distributions are derived in section 5.4 for W-QoS NOMA followed by sum rate results for the ideal case where cluster pairs have perfectly correlated channels. Numerical results are presented in section 5.4.4 and conclusions are drawn in section 5.5.

5.2 NOMA-BF system model

We consider a multiuser BF system in which the BS is equipped with M_T antennas. The BS supports Ω UEs selected from L possible candidates. We assume that all UEs are equipped with a single antenna and $L \geq \Omega$. To integrate NOMA with a multiuser system scenario, the Ω users are grouped into N_1 clusters with B highly correlated users in each cluster and N_2 singletons, i.e., $\Omega = N_1 B + N_2$. The basic NOMA-BF

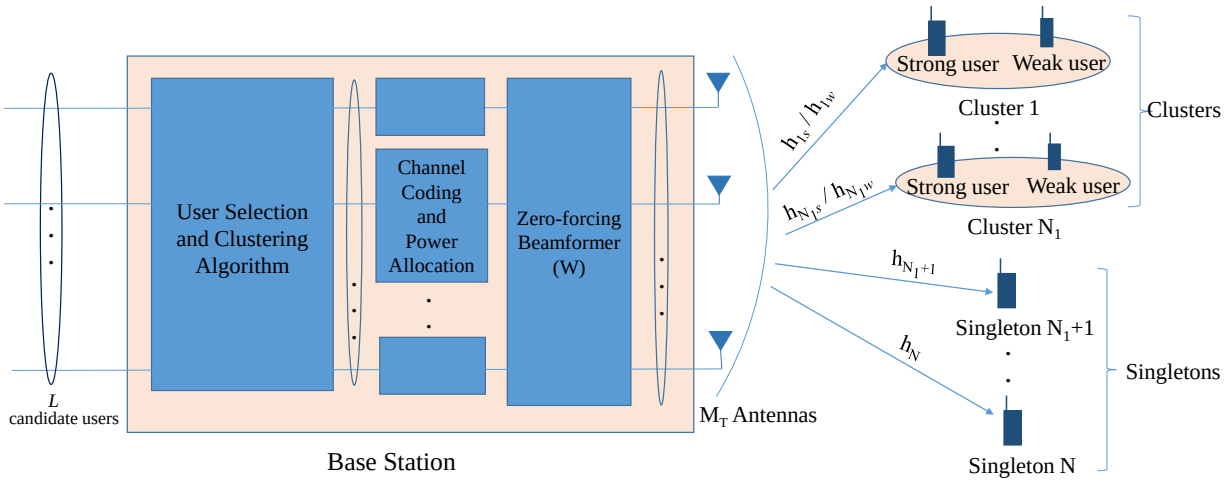


Figure 5-1: NOMA-BF downlink system with user selection and clustering algorithm.

system with $B = 2$ users in each of the N_1 clusters is illustrated in Fig. 5-1. ZF is used to separate the clusters and singletons. Within clusters, the strong and weak users are separated using NOMA power allocation and SIC.

Let $N = N_1 + N_2$, where $M_T \geq N$, then the $N \times M_T$ Rayleigh fading channel matrix from the BS to the strong users in the N_1 clusters and the N_2 singletons is given by

$$\mathbf{H} = [\underbrace{\mathbf{h}_{1s}^T, \mathbf{h}_{2s}^T, \cdots, \mathbf{h}_{N_1 s}^T}_{\text{Clusters}}, \underbrace{\mathbf{h}_{N_1+1}^T, \cdots, \mathbf{h}_N^T}_{\text{Singletons}}]^T. \tag{5.1}$$

The strengths of the N channels in (5.1) are defined by $E|(\mathbf{h}_{is})_r|^2 = \beta_{is}$, where $(\mathbf{h}_{is})_r$ is the r^{th} element of $\mathbf{h}_{is}$, and $E|(\mathbf{h}_i)_r|^2 = \beta_i$. Therefore, $\mathbf{H}$ in (5.1) can be written as

$$\mathbf{H} = \text{diag}\left(\sqrt{\beta_{1s}}, \cdots, \sqrt{\beta_{N_1 s}}, \sqrt{\beta_{N_1+1}}, \cdots, \sqrt{\beta_N}\right) \mathbf{G}, \tag{5.2}$$

where the $N \times M_T$ matrix $\mathbf{G}$ contains and I.I.D. $\mathcal{CN}(0,1)$ entries. Throughout, we use subscripts s and w to denote the strong/near and weak/far users in the N_1 clusters, respectively. Moreover, $\mathbf{h}_{is}$ represents the channel vector of the strong user in the i^{th} cluster and $\mathbf{h}_{N_1+i}$ is the channel vector of the i^{th} singleton. Similarly, for the weak users, $\mathbf{h}_{iw}$ is the channel vector of the weak user in the i^{th} cluster where $E|(\mathbf{h}_{iw})_r|^2 = \beta_{iw}$. The transmitter performs ZF based on $\mathbf{H}$ and the design of the

$M_T \times N$ un-normalized precoding matrix is given by[1]

$$\mathbf{W}_0 = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1} = [\mathbf{w}_{01}, \mathbf{w}_{02}, \cdots, \mathbf{w}_{0N}], \tag{5.3}$$

so that the normalized precoder is

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N], \tag{5.4}$$

where $\mathbf{w}_i = \mathbf{w}_{0i}/||\mathbf{w}_{0i}||$. Now, the $N \times 1$ received signal vector at the strong users and singletons can be expressed as

$$\mathbf{r} = \mathbf{H}\mathbf{W}\mathbf{x} + \mathbf{z}, \tag{5.5}$$

where $\mathbf{x}$ is the signal to be transmitted and

$$\mathbf{z} = [z_{1s}, \ldots, z_{N_1 s}, z_{N_1+1}, \ldots, z_N]^T \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N) \tag{5.6}$$

denotes the noise vector. Further, $\mathbf{x}$ can be represented as

$$\mathbf{x} = [x_1, \cdots, x_{N_1}, x_{N_1+1}, \cdots, x_N]^T, \tag{5.7}$$

where $x_1 \cdots x_{N_1}$ are the signals for the clusters given by

$$x_i = \left(x_{is}\alpha_i + x_{iw}\sqrt{1-\alpha_i^2}\right), \tag{5.8}$$

and $x_{N_1+1} \cdots x_N$ are the signals for singletons. Further, x_{is} and x_{iw} are the strong user signal and weak user signal for cluster i, respectively. α_i and $\sqrt{1-\alpha_i^2}$ are the amplitude scalings and α_i^2 and $1-\alpha_i^2$ are the power fractions of x_{is} and x_{iw}, respectively. Similarly, the $N_1 \times 1$ received signal vector at the weak users is defined

[1] As discussed in Sec. 2.5, the main concern on the performance of ZFBF is that the performance degrades as the cross correlation between the user beams increases. However, in NOMA, only the strong users channel is considered in a cluster during precoder design to reduce the inter-stream interference.

as

$$\mathbf{r}_w = \mathbf{H}_w \mathbf{W}\mathbf{x} + \mathbf{z}_w, \tag{5.9}$$

where

$$\mathbf{H}_w = [\mathbf{h}_{1w}, \ldots, \mathbf{h}_{N_1 w}] \tag{5.10}$$

and

$$\mathbf{z}_w = [z_{1w}, \ldots, z_{N_1 w}]^T \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_1}). \tag{5.11}$$

In this chapter, we assume that $\sigma^2 = 1$.

In Sec. 5.2.1 - Sec. 5.2.4, we provide the background results and notation required for the novel algorithm and performance analysis developed in Secs. 5.2.5, 5.3 and 5.4, respectively.

5.2.1 SNR of a singleton

NOMA-singletons are the users that cannot be grouped into a highly correlated pair. The received signal for the i^{th} singleton can be written as

$$r_i = \mathbf{h}_i \mathbf{W}\mathbf{x} + z_i = \frac{x_i}{||\mathbf{w}_{0i}||} + z_i, \tag{5.12}$$

for $i = N_1 + 1, \ldots, N$. Further, we have [33]

$$\frac{1}{||\mathbf{w}_{0i}||^2} = \frac{1}{[(\mathbf{H}\mathbf{H}^H)^{-1}]_{ii}} \triangleq \frac{\beta_i}{[(\mathbf{G}\mathbf{G}^H)^{-1}]_{ii}}. \tag{5.13}$$

Also we note that

$$\left[(\mathbf{G}\mathbf{G}^H)^{-1}{}_{ii}\right]^{-1} = Y_i \sim \chi_p^2, \tag{5.14}$$

where $p = M - N + 1$ and χ_p^2 represents a complex chi-squared variable with p degrees of freedom. Now (5.12) can be rewritten as

$$r_i = \sqrt{\beta_i Y_i} x_i + z_i. \tag{5.15}$$

The received SNR of singleton i is then

$$\text{SNR}_i = \frac{\beta_i Y_i E_x}{\sigma^2} = \beta_i Y_i \rho, \quad i = N_1 + 1, \ldots, N, \tag{5.16}$$

where the signal strength is $E_x = E|x_i|^2$ and the transmit SNR is $\rho = \frac{E_x}{\sigma^2}$. Hence, the rate of a singleton is given by

$$R_i = \log(1 + \beta_i Y_i \rho), \tag{5.17}$$

for $i = N_1 + 1, \ldots, N$.

5.2.2 SNR of the strong user in a cluster

In NOMA clusters, both strong and weak users are served by the BS using the same time and frequency, but with different power levels. The strong user is served with less power as it has better channel conditions than the weak user. The ZF precoders are designed based on the strong users' channels as derived in (5.1) - (5.4). Thus, from (5.5) and (5.13), the received signal for the strong user in the i^{th} cluster is given by

$$\begin{aligned} r_{is} = \ & \mathbf{h}_{is}\mathbf{W}\mathbf{x} + z_{is} = \frac{x_i}{||\mathbf{w}_{0i}||} + z_{is}, \\ = & \left(x_{is}\alpha_i + x_{iw}\sqrt{1-\alpha_i^2}\right)\sqrt{\beta_{is}Y_i} + z_{is}. \end{aligned} \tag{5.18}$$

SIC is used to first detect the weak user's signal, then that of the strong user. During SIC, at the strong user receiver, it must be ensured that the detected signal of the weak user has no error propagation. When detecting x_{iw}, from (5.18), the SINR of the weak user at the strong user is given by

$$\text{SINR}_{is}^{(SIC)} = \frac{\beta_{is}Y_i(1-\alpha_i^2)E_x}{\beta_{is}Y_i\alpha_i^2 E_x + \sigma^2}, \; i = 1, 2, \ldots, N_1. \tag{5.19}$$

If there is no error propagation during the detection of x_{iw}, i.e., after successful SIC of the weak user signal, then, from (5.18), the SNR of the strong user is given by

$$\mathrm{SNR}_{is} = \beta_{is} Y_i \alpha_i^2 \rho, \quad i = 1, 2, \ldots, N_1. \tag{5.20}$$

5.2.3 SINR of the weak user in a cluster

Similarly, the received signal for a weak user in a cluster is given by

$$r_{iw} = \mathbf{h}_{iw}\mathbf{w}_i x_i + \mathbf{h}_{iw} \sum_{j=1, j\neq i}^{N} \mathbf{w}_j x_j + z_{iw}. \tag{5.21}$$

Since ZF is only performed with the strong users' channels, $\mathbf{h}_{iw}\mathbf{w}_j \neq 0$ for $j \neq i$ and the weak user suffers from interference from other clusters and from their strong user partner. Hence, the SINR of the weak user is given by

$$\mathrm{SINR}_{iw} = \frac{|\mathbf{h}_{iw}\mathbf{w}_i|^2(1-\alpha_i^2)\rho}{|\mathbf{h}_{iw}\mathbf{w}_i|^2\alpha_i^2\rho + \sum_{j\neq i} |\mathbf{h}_{iw}\mathbf{w}_j|^2\rho + 1}, \quad i = 1, 2, \ldots, N_1. \tag{5.22}$$

5.2.4 Cluster rate

In NOMA, the total transmit power of a BF vector is divided between the weak and strong users in order to enhance performance and often to provide gains relative to a standard TDMA implementation. For example, the power allocation scheme proposed in [40] maximizes the sum rate of the system while keeping the weak user's rate greater or equal to that of a comparable TDMA multiuser BF system. Similarly, [96] proposed to maximize the sum rate of both strong and weak users while maintaining a rate greater than or equal to that of a comparable TDMA system. To enable a comparison with TDMA, we note that TDMA-BF with M_T transmit antennas requires 2 time slots to accommodate $2M$ users [40] [96]. However, with NOMA, $2M$ users can be accommodated per time slot. If $R_{s,\mathrm{TDMA}}$ and $R_{w,\mathrm{TDMA}}$ are the rates of strong and weak users supported by a conventional TDMA system then the comparable rate for

this pair is

$$R_{\text{TDMA}} = \frac{1}{2} R_{s,\text{TDMA}} + \frac{1}{2} R_{w,\text{TDMA}}. \tag{5.23}$$

Further, the NOMA cluster rate is given by

$$R_{\text{NOMA}} = \log(1 + \text{SNR}_{is}) + \log(1 + \text{SINR}_{iw}). \tag{5.24}$$

In our NOMA implementation we ensure $R_{\text{NOMA}} > R_{\text{TDMA}}$ for all selected clusters as described in the following.

5.2.5 W-QoS based NOMA (W-QoS NOMA) algorithm

From Sec. 5.2.2, 5.2.3 and 5.2.4, we see that the basic requirements of power allocation in NOMA are to achieve successful SIC, acceptable strong and weak user rates and a high sum rate exceeding that of conventional TDMA. In this section, we outline a novel approach to achieving these objectives, namely the W-QoS NOMA approach. In this approach, we assume that the weak user's SINR during weak user signal detection is always smaller than the weak user's SINR at the strong user. Under this assumption, if a target SINR satisfies the weak user detection, then it will be high enough for successful SIC detection at the strong user.

In brief, the proposed W-QoS NOMA technique first selects the smallest SINR threshold that the weak user can accept for both individual rate and successful SIC requirements, i.e., $\text{SINR}_{iw} = T$ with $T \geq 1$[2]. Then, the power is allocated to the strong user as per (5.26) below. Similarly, the power scaling factor for the weak user can be calculated using $1 - \alpha_i^2$. Using the above parameters, if $R_{\text{NOMA}} > R_{\text{TDMA}}$, then a cluster is possible that exceeds the TDMA rate. This is the largest possible cluster rate that exceeds the TDMA rate while satisfying the SIC constraint at the strong user and providing a guaranteed SINR to the weak user. If one of the above conditions is not satisfied then the cluster will be dropped in favor of the strong user supported as a singleton. The W-QoS NOMA algorithm is outlined below:

[2]The requirement that $T \geq 1$ is derived in Sec. 5.4.1.

1. Employ a user selection and clustering algorithm to select Ω UEs and N_1 clusters (see Sec. 5.2.6).

2. $i = 1$.

3. Set $\text{SINR}_{iw} = T$ for cluster i[3], from (5.22), giving

$$\text{SINR}_{iw} = \frac{|\mathbf{h}_{iw}\mathbf{w}_i|^2(1-\alpha_i^2)\rho}{|\mathbf{h}_{iw}\mathbf{w}_i|^2\alpha_i^2\rho + \sum\limits_{j\neq i}|\mathbf{h}_{iw}\mathbf{w}_j|^2\rho + 1} = T. \tag{5.25}$$

4. Solve (5.25) to calculate

$$\alpha_i = \sqrt{\frac{|\mathbf{h}_{iw}\mathbf{w}_i|^2\rho - T\Big(\sum\limits_{j\neq i}|\mathbf{h}_{iw}\mathbf{w}_j|^2\rho + 1\Big)}{|\mathbf{h}_{iw}\mathbf{w}_i|^2\rho(T+1)}}. \tag{5.26}$$

5. If $|\mathbf{h}_{iw}\mathbf{w}_i|^2\rho > T\Big(\sum\limits_{j\neq i}|\mathbf{h}_{iw}\mathbf{w}_j|^2\rho + 1\Big)$, then $0 < \alpha_i < 1$ and the cluster remains possible.

6. Check whether the cluster rate exceeds the TDMA rate, where NOMA and TDMA cluster rates are given by

$$R_{\text{NOMA}} = \log(1 + |\mathbf{h}_{is}\mathbf{w}_i|^2\alpha_i^2\rho) + \log(1+T), \tag{5.27}$$

and

$$R_{\text{TDMA}} = \frac{1}{2}\log(1 + |\mathbf{h}_{is}\mathbf{w}_i|^2\rho) + \frac{1}{2}\log(1 + |\mathbf{h}_{iw}\mathbf{w}_{iw}|^2\rho), \tag{5.28}$$

respectively. Note that $\mathbf{w}_{iw}$ is the i^{th} column of the precoding matrix when the weak user rather than the strong user in cluster i is served, ie, $\mathbf{h}_{is}$ is removed from the channel matrix, $\mathbf{H}$, and replaced by $\mathbf{h}_{iw}$. To obtain a NOMA cluster

[3]This also implies that $\text{SINR}_{is}^{(SIC)} \geq T$ as shown in Sec. 5.4.

rate that exceeds the TDMA rate, we require

$$\log(1+|\mathbf{h}_{is}\mathbf{w}_i|^2\alpha_i^2\rho)+\log(1+T)>\frac{1}{2}\Big(\log(1+|\mathbf{h}_{is}\mathbf{w}_i|^2\rho)+\log(1+|\mathbf{h}_{iw}\mathbf{w}_{iw}|^2\rho)\Big). \tag{5.29}$$

Eliminating the log term, we get

$$(1+T)(1+|\mathbf{h}_{is}\mathbf{w}_i|^2\alpha_i^2\rho)>\sqrt{(1+|\mathbf{h}_{is}\mathbf{w}_i|^2\rho)(1+|\mathbf{h}_{iw}\mathbf{w}_{iw}|^2\rho)}. \tag{5.30}$$

For notational simplicity, let A be equal to the right hand side of (5.30) which can be rewritten as

$$\alpha_i>(A/(1+T)-1)^{1/2}\left(|\mathbf{h}_{is}\mathbf{w}_i|^2\rho\right)^{-1/2}. \tag{5.31}$$

If (5.31) is satisfied, then the cluster rate exceeds the TDMA rate.

7. If steps 4 and 5 are satisfied, then the NOMA cluster is accepted. If either step 4 or 5 fail, then the cluster is rejected in favor of a singleton containing the strong user.

8. Let $i=i+1$, go to step 3 and repeat with the next cluster until N ZF beams are created.

Note that the computational complexity of W-QoS NOMA is almost identical to the approaches in [40, 96] and has lower complexity to that in [111] which requires non-convex optimization. Beyond the well known selection and clustering approach in step 1 of the algorithm, the computational complexity comparison involves the cluster acceptance procedure in steps 2 - 8 which are iterated N times. This requires computation of $\mathbf{h}_{iw}\mathbf{w}_j$, $j=1,2,\dots N$ and $\mathbf{h}_{is}\mathbf{w}_i$ which involves $(N+1)M_T$ complex multiplications. The vector $\mathbf{w}_{iw}$ also requires recomputation of $\mathbf{W}$ requiring a matrix inversion of $\mathcal{O}(N^3)$ and $2NM_T$ complex multiplications. Finally, step 5 requires a logical comparison and calculation of equations (5.26), (5.27), (5.28) and (5.31), where the components of the equations are scalars as the vector products have already been

computed. These calculations are dominated by the $\mathcal{O}(N^3)$ inverse which gives the total complexity order as $\mathcal{O}(N^4)$ due to the iteration.

The rationale behind the proposed algorithm is to guarantee the weak users QoS and still achieve the NOMA rate higher than TDMA rate. Moreover, W-QoS NOMA technique is very simple and easy to merge with established user selection and clustering algorithms. Further, the power allocation is optimal under the given constraint in W-QoS NOMA.

5.2.6 Semi-correlated user clustering algorithms

From (5.22), we see that imperfect correlation between the strong and weak user channels reduces the desired signal power (ie, reduces $|\mathbf{h}_{iw}\mathbf{w}_i|^2$) and increases the interference (ie, increases $\sum_{j\neq i} |\mathbf{h}_{iw}\mathbf{w}_j|^2$), hence lowering SINR_{iw}. Furthermore, reduced orthogonality among the strong user and singleton channels lowers ZF performance. Hence, the ideal candidates for W-QoS NOMA are users with highly correlated strong and weak user channels and near orthogonality among the strong users and singletons. In order to schedule users with these properties we combine the CUC [40] and SUS-UM [96] methods as described below.

In CUC, correlated strong and weak user channels are deliberately selected by setting a correlation threshold, Ψ_0, so that only users with a correlation greater than Ψ_0 are considered. Furthermore, from the set of highly correlated users, the pair with the largest power difference is selected.[4] However, CUC does not directly select highly orthogonal strong user and singleton channels. In contrast, the SUS-UM technique [96], initially selects a set of semi-orthogonal users using the SUS algorithm [33] which only considers users which satisfy an orthogonality rule (defined by the threshold value, Γ). This is followed by user matching via a rate maximization technique. However, SUS-UM does not directly select highly correlated strong and weak users. Therefore, for implementation with W-QoS NOMA, we merge the two approaches, so that the initial SUS algorithm in SUS-UM is followed by the user pairing approach in CUC. This combined algorithm is denoted SUS-CUC. Hence, we

[4]The relevance of a power differential is discussed in more detail in Sec. 5.4.4.

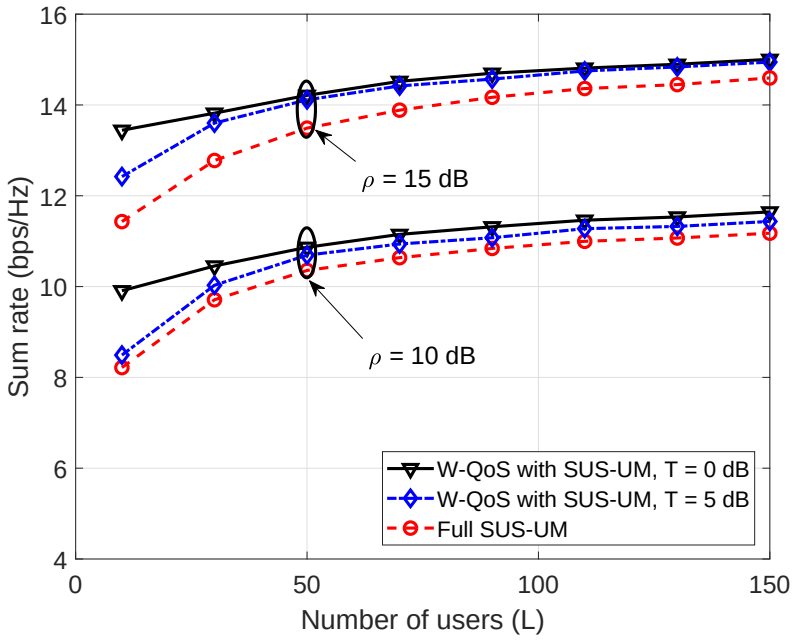


Figure 5-2: Sum rate versus the number of users. A performance comparison of SUS-UM based W-QoS NOMA with the full SUS-UM technique. $M = 2$, $\sigma^2 = 1$, $\Gamma = 0.4$ and N (N_1 and N_2) depend upon the UM technique.

have two potential user selection and clustering techniques which attempt to create near-orthogonal clusters: SUS-UM from [96] and SUS-CUC, our merger of [40] and [96]. Both selection algorithms are integrated with the W-QoS NOMA technique to evaluate the performance in Sec. 5.3.

5.3 Performance comparison

In this section, we demonstrate the performance of the W-QoS NOMA algorithm with SUS-UM and SUS-CUC. For simplicity, we assume normalized independent Rayleigh fading channels for all users so that $\beta_{is} = \beta_{iw} = \beta_i = 1$ in order to compare SUS-UM with SUS-CUC. Later on, in Sections 5.4 and 5.4.4 we present the implications of power variation across the users.

First, we follow the full user matching and power allocation technique used in [96], referred to here as full SUS-UM, and compare it with our proposed W-QoS NOMA technique using the clusters selected by SUS-UM. From Figs. 5-2 and 5-3, we observe

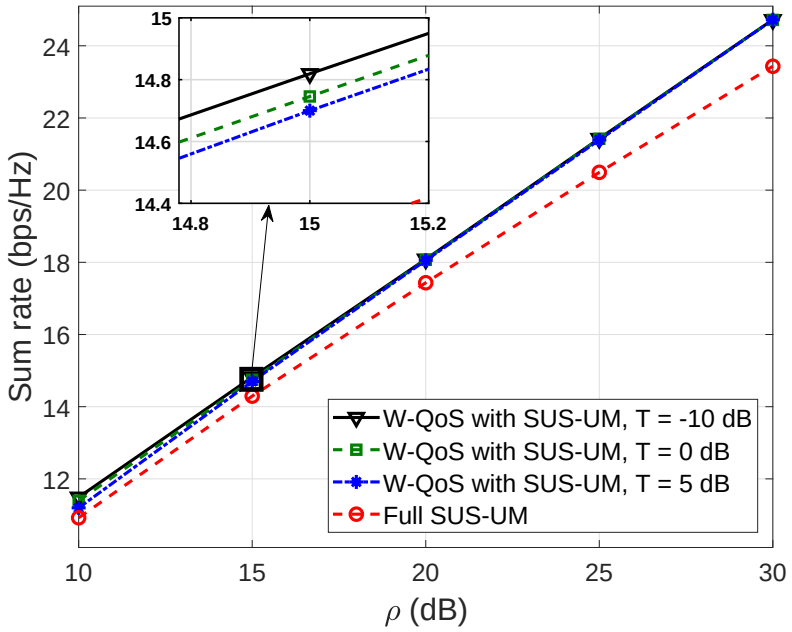


Figure 5-3: Sum rate versus SNR (ρ). A performance comparison of SUS-UM based W-QoS NOMA with the full SUS-UM technique. $M = 2$, $L = 100$, $\sigma^2 = 1$, $\Gamma = 0.4$ and N (N_1 and N_2) depend upon the UM technique.

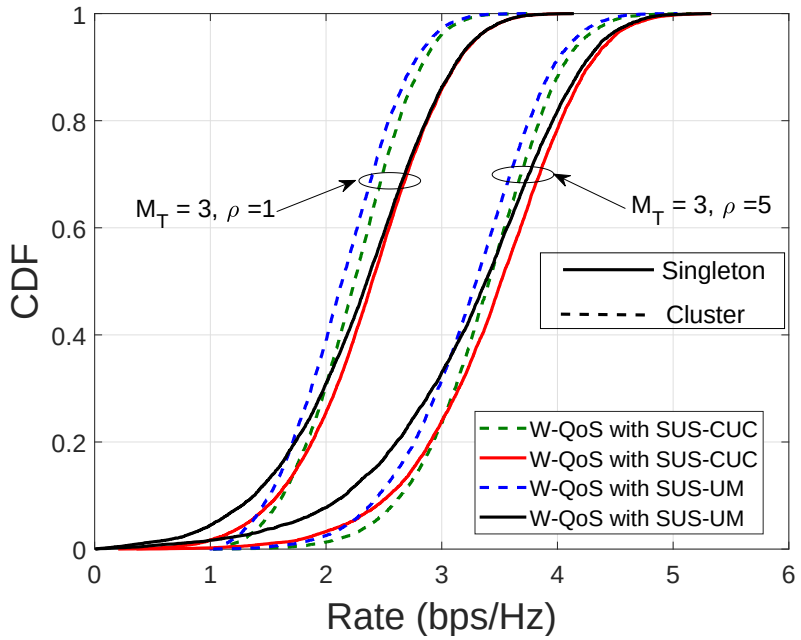


Figure 5-4: Rate CDFs of clusters and singletons for $L = 10$. For SUS-CUC, $T = 1$, $\Gamma = 0.4$ and $\Psi_0 = 0.75$ and for SUS-UM, $\Gamma = 0.4$. $N(N_1$ and $N_2)$ depend upon Ψ_0 and Γ.

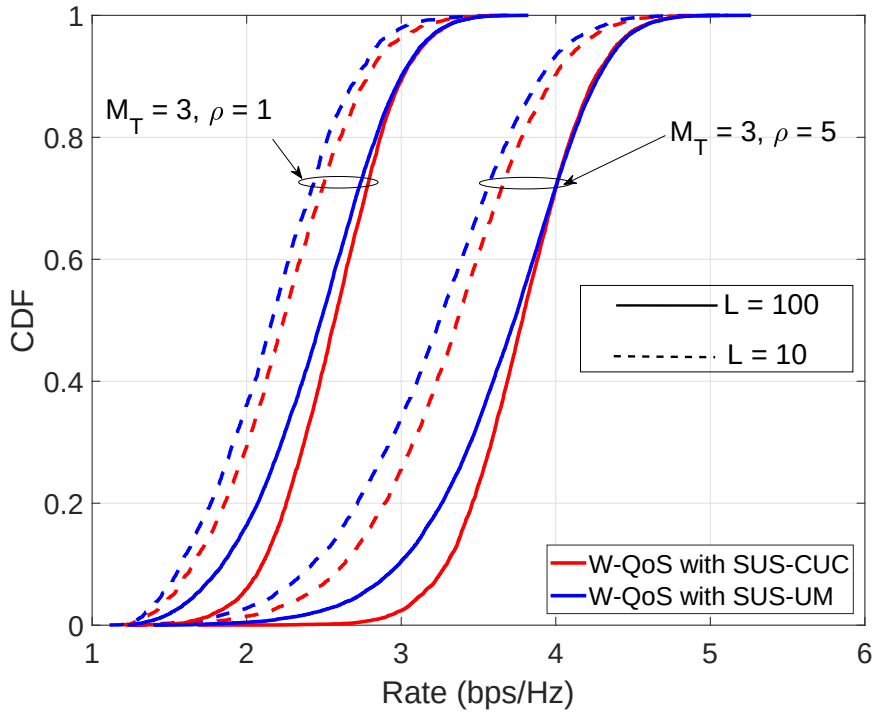


Figure 5-5: Rate CDFs of clusters for $L = 10$ and 100. For SUS-CUC, $T = 1$, $\Gamma = 0.4$ and $\Psi_0 = 0.75$ and for SUS-UM, $\Gamma = 0.4$. $N(N_1$ and $N_2)$ depend upon Ψ_0 and Γ.

that SUS-UM based W-QoS NOMA has a performance gain over the full SUS-UM. The performance gain is due to the weak user QoS requirement which makes sure that the system allocates the minimal, but required power fraction to the weak user. This is verified during simulations where we observe that in the full SUS-UM approach the system can give more power to the weak users to compensate for the poorer channel conditions. The effect of using the target rate is also shown in Figs. 5-2 and 5-3. In Fig. 5-2, we observe that the performance of W-QoS NOMA with SUS-UM increases as T gets smaller. However, the effect of T diminishes with an increasing number of users, L. Moreover, the effect of T is insensitive to ρ which is clearly shown in Fig. 5-3.

Second, to demonstrate the importance of the correlation between clustered user channels, in Figs. 5-4 and 5-5 we compare the cluster and singleton rates achieved by both SUS-CUC and SUS-UM techniques with W-QoS NOMA. As expected, the performance of both techniques increases with an increase in transmit SNR, ρ, and

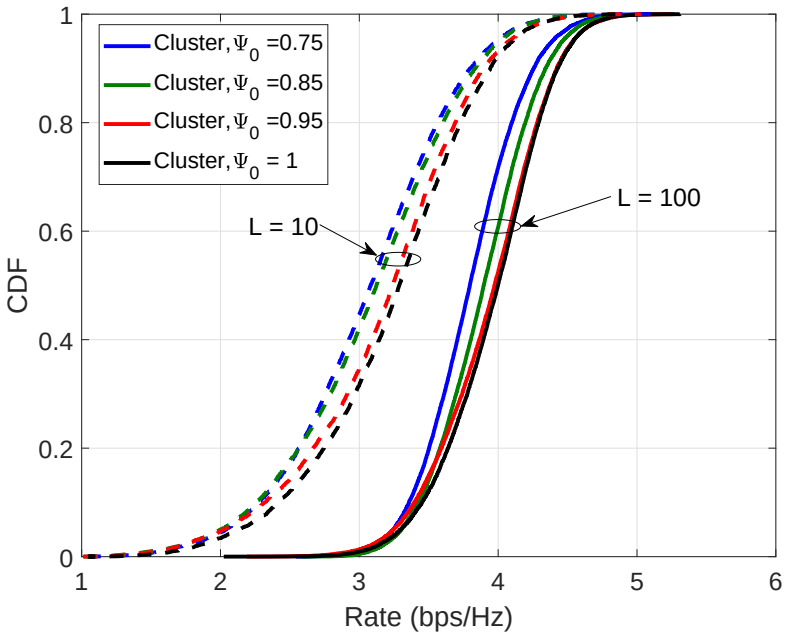


Figure 5-6: Rate CDFs of clusters for SUS-CUC algorithm for $\Psi_0 = 0.75$, 0.95 and 1 (perfect correlation). $\Gamma = 0.4$, $\rho = 5$ dB, $T = 1$, $M_T = 2$. $N(N_1$ and $N_2)$ depend upon Ψ_0 and Γ, where $N \leq M_T$.

candidate users, L. Further, SUS-CUC has a marginal performance gain over SUS-UM. This is because in SUS-CUC, clusters are only formed if the user channels exceed a correlation threshold. In contrast, SUS-UM only considers channel strengths during cluster pair selection. Note that the initial SUS selection approach is the same for both techniques, but simulations show that the paired users are different.

In Fig. 5-6 we compare W-QoS NOMA based SUS-CUC for $\Psi_0 = 0.75$, 0.95 and 1 (perfect correlation). In the simulations, we use $\Gamma = 0.4$, $M_T = 2$, while N_1 and N_2 depend upon Ψ_0 and Γ, where $N \leq M_T$. The observation in Fig. 5-6 is consistent with the observation made in [40] that high correlation reduces the ICI which further enhances the weak user performance. Here, we can observe that the cluster rate with $\Psi_0 = 0.95$ is higher than that with $\Psi_0 = 0.75$. In addition, when the number of users increases, as shown in Fig. 5-6, the performance of the system increases due to multiuser diversity gain. In order to obtain the benchmark performance of perfectly correlated users (which doesn't occur via SUS as no randomly generated channels are perfectly correlated) we set the weak user channels as $\mathbf{h}_{iw} = \mathbf{h}_{is} \times \frac{||\mathbf{h}_{iw}||}{||\mathbf{h}_{is}||}$. Fig. 5-6

illustrates that cluster performance increases with an increase in correlation between the clustered users and $\Psi_0 = 1$ provides an upper bound on cluster rate. Although $\Psi_0 = 1$ is unrealistic, it is useful to understand how the correlation factor affects the cluster rate and how close partially correlated channels can get to the perfectly correlated bound.

Three traits can be observed from Sec. 5.3. First, the advantage of the proposed W-QoS NOMA based power allocation technique, i.e., W-QoS NOMA with SUS-UM has a performance gain over full SUS-UM. Secondly, despite the fact that both SUS-UM and SUS-CUC techniques use the same SUS based semi-orthogonal user selection approach, SUS-CUC has a marginal performance gain due to its clustering approach. The third observation is that the correlation factor affects the cluster rate and we observe that high correlation reduces the ICI which further enhances the weak user performance and ultimately increases the cluster rate.

5.4 Analysis

As observed in Sec. 5.3, the performance gain of the NOMA system is dependent on the correlation between the users in a cluster. Thus, in this section, we analyze the SINR and the sum rate of W-QoS NOMA for arbitrary correlation and provide closed form upper bounds on performance based on perfect correlation.

If $\mathbf{h}_{iw}$ and $\mathbf{h}_{is}$ are correlated with correlation factor, Ψ_i, then the channels are related by

$$\mathbf{h}_{iw} = \Psi_i \sqrt{\frac{\beta_{iw}}{\beta_{is}}} \mathbf{h}_{is} + \sqrt{\beta_{iw}} \sqrt{1 - \Psi_i^2} \mathbf{e}_i, \tag{5.32}$$

where $\mathbf{e}_i$ is an I.I.D. $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{M_T})$ vector. Since ZF is performed on the strong user channels we have $\mathbf{h}_{is}\mathbf{w}_i = \sqrt{\beta_{is} Y_i}$ and $|\mathbf{h}_{is}\mathbf{w}_j| = 0$ for $j \neq i$. Hence, from (5.22) and (5.32) we obtain

$$\text{SINR}_{iw} = \frac{|\Psi_i \sqrt{Y_i} + \sqrt{1 - \Psi_i^2} P_{ii}|^2 (1 - \alpha_i^2) \rho}{|\Psi_i \sqrt{Y_i} + \sqrt{1 - \Psi_i^2} P_{ii}|^2 \alpha_i^2 \rho + \sum_{j \neq i} \rho (1 - \Psi_i^2) |P_{ij}|^2 + \beta_{iw}^{-1}}, \tag{5.33}$$

where $P_{ii} = \mathbf{e}_i \mathbf{w}_k$. Simplifying (5.33) gives

$$\text{SINR}_{iw} = \frac{\beta_{iw} Y_i (1-\alpha_i^2)\rho\Delta_i}{\beta_{iw} Y_i \alpha_i^2 \rho \Delta_i + \sum_{j \neq i} \beta_{iw}\rho(1-\Psi_i^2)|P_{ij}|^2 + 1}, \tag{5.34}$$

where $\Delta_i = |\Psi_i + \sqrt{(1-\Psi_i^2)/Y_i} P_{ii}|^2$. Note that when the strong and weak user are perfectly correlated, $\Psi = 1$, then $\Delta_i = 1$ and $1 - \Psi_i^2 = 0$. Hence, when $\Psi = 1$ we have the simplified result

$$\text{SINR}_{iw} = \frac{\beta_{iw} Y_i (1-\alpha_i^2)\rho}{\beta_{iw} Y_i \alpha_i^2 \rho + 1}. \tag{5.35}$$

Comparing (5.34) and (5.35), we observe that the imperfect correlation introduces the power penalty, Δ_i, and the interference term $\sum_{j \neq i} \beta_{iw}\rho(1-\Psi_i^2)|P_{ij}|^2$.

Before considering any further analysis, we first draw some insights from the structure of (5.34). First, imperfect correlation creates the interference term in the denominator of (5.34) which has mean value $(N-1)\beta_{iw}\rho(1-\Psi_i^2)$. Hence, interference grows linearly with the weak user channel power, β_{iw}, the dimension, N, and the SNR, ρ, while it decreases proportional to $1-\Psi_i^2$. Secondly, the mean of the power penalty term, Δ_i, is $E[\Delta_i] = \Psi_i^2 + [1-\Psi_i^2]E[Y_i^{-1}]$. This follows from the fact that the P_{ij} variables are $\mathcal{CN}(0,1)$ since the $\mathbf{w}_j$ vectors are unit norm. The inverse moment, $E[Y_i^{-1}]$, is well known [33], so we have $E[\Delta_i] = \Psi_i^2 + [1-\Psi_i^2]/(M_T - N) < 1$ for $M_T > N$. Hence, the power penalty increases with N and grows quadratically with $|\Psi_i|$. This quantifies how the weak user SINR is affected by the system parameters.

For the strong user we note that as long as $\text{SINR}_{iw} = T$ is satisfied, both cluster formation and rate are helped by large α_i (see (5.31) and (5.20) respectively). Rearranging (5.33) and computing α_i^2 from $\text{SINR}_{iw} = T$, we get

$$\alpha_i^2 = \frac{1 - T\left\{ \dfrac{(1-\Psi_i^2)\sum_{j \neq i} |P_{ij}|^2 + \dfrac{1}{\rho\beta_{iw}}}{|\Psi_i\sqrt{Y_i} + \sqrt{1-|\Psi|^2}P_{ii}|^2} \right\}}{1+T}. \tag{5.36}$$

Note that α_i^2 is an increasing function of Ψ, ρ and β_{iw}. Hence both rate and cluster formation are improved by large values of Ψ, ρ and β_{iw}. As expected, performance improves with Ψ and ρ, but it is interesting that W-QoS NOMA does not need a channel power differential between strong and weak users. In fact, performance is improved when both channels are strong[5].

Finally, we note that for large values of Y_i, $\alpha_i^2 \approx (1+T)^{-1}$ from (5.36) and $\text{SNR}_{is} \approx \beta_{is} Y_i \rho (1+T)^{-1}$. Similarly, with perfect correlation, $\text{SNR}_{is} \approx \beta_{is} Y_i \rho (1+T)^{-1}$ from (5.20) and (5.36). Hence, the SNR is similar for the perfectly correlated case and the imperfectly correlated case for large values of Y_i. Hence, the upper tail of the CDFs is similar which supports the tightness of the upper bound shown in Sec. 5.4.4.

Although the P_{ij}'s have simple distributions ($\mathcal{CN}(0,1)$), they are correlated due to their dependence on the $\mathbf{w}_j$'s. As a result, SINR_{iw} in (5.33) and (5.34), is a complicated non-linear function of several random variables involving powers and ratios of chi-squared variables, Y_i, and correlated Gaussians, P_{ij}. It seems that analytical progress based on (5.33) and (5.34) is extremely difficult and so we restrict our analysis to the ideal case where $\Psi_0 = 1$ giving an upper bound on performance (see section 5.3). Therefore, in Secs. 5.4.1 and 5.4.3 we focus on perfectly correlated strong and weak user channels where ZF also removes ICI for the weak users, i.e., $|\mathbf{h}_{iw}\mathbf{w}_j| = 0$ for $j \neq i$. In this scenario, SINR_{iw} is given by (5.35) and some straightforward algebra shows that the SINR in (5.35) is less than $\text{SINR}_{is}^{(SIC)}$ in (5.19). This proves that the W-QoS NOMA algorithm which sets $\text{SINR}_{iw} = T$ also satisfies $\text{SINR}_{is}^{(SIC)} \geq T$.

5.4.1 SIC detection at strong user

As discussed in Secs. 5.2.2 and 5.2.5, the first basic requirement of power allocation in NOMA is to achieve successful SIC. Therefore, in this section, we analyze a condition that enables successful detection and cancellation of the weak user signal at the strong user. To achieve this, we assume perfectly correlated strong and weak user channels in a cluster.

[5]In Sec. 5.4.4, we use the power differential in SUS-CUC to investigate the rate dependence on channel powers in practical clusters.

When $\Psi_i = 1$, step (3) of the W-QoS NOMA algorithm gives

$$\mathrm{SINR}_{iw} = \frac{\beta_{iw} Y_i (1 - \alpha_i^2)}{\beta_{iw} Y_i \alpha_i^2 + \frac{1}{\rho}} = T. \tag{5.37}$$

Solving (5.37) gives the strong user power scaling factor as

$$\alpha_i^2 = \frac{\beta_{iw} Y_i - \frac{T}{\rho}}{\beta_{iw} Y_i (T+1)}, \tag{5.38}$$

and the resulting SINR_{iw} is fixed at T. Consequently, the weak user power scaling factor can be derived using $(1 - \alpha_i^2)$. Substituting (5.38) into (6.6) gives

$$\mathrm{SNR}_{is} = \frac{\beta_{is} \rho \Big(\beta_{iw} Y_i - \frac{T}{\rho} \Big)}{\beta_{iw} (T+1)}. \tag{5.39}$$

From (5.38) and (5.39), we require $\beta_{iw} Y_i \geq \frac{T}{\rho}$. Let us suppose that $\beta_{iw} Y_i = \frac{T}{\rho} + \zeta$, then (5.38) becomes

$$\alpha_i^2 = \frac{\zeta}{\Big(\frac{T}{\rho} + \zeta \Big) \Big(T + 1 \Big)}. \tag{5.40}$$

It is clear that we require $0 < \alpha_i^2 < 1$. Further, the weak user power, $(1 - \alpha_i^2)$, is to be greater than the strong user power, α_i^2. Therefore, $1 - \alpha_i^2 > \alpha_i^2$, which yields $\alpha_i^2 < 1/2$. Thus,

$$\alpha_i^2 = \frac{\zeta}{\Big(\frac{T}{\rho} + \zeta \Big)} \frac{1}{\Big(T + 1 \Big)} \leq \frac{1}{T+1} \leq \frac{1}{2}. \tag{5.41}$$

Therefore, $T \geq 1$ is a sufficient condition for the threshold SINR to enable a successful power allocation.

5.4.2 Cluster rate maximization

The next basic requirement in NOMA is that the NOMA cluster rate should always be higher then the conventional TDMA rate as discussed in Sec. 5.2.4. Therefore, in this section, we analyze the condition required to obtain $R_{\mathrm{NOMA}} > R_{\mathrm{TDMA}}$.

The total NOMA cluster rate is given in (5.24). Therefore, substituting (5.37)

and (5.39) into (5.24) gives

$$R_{\text{NOMA}} = \log\Big(1 + \rho\beta_{is}Y_i - T\Big(\frac{\beta_{is}}{\beta_{iw}} - 1\Big)\Big). \tag{5.42}$$

Since, $\frac{\beta_{is}}{\beta_{iw}} \geq 1$, it follows that cluster rates are maximized by using the minimum possible T. Hence, T in (5.37) is the smallest possible target to satisfy the weak user. Therefore, from (5.29), to obtain the NOMA cluster that exceeds the TDMA rate, we need $R_{\text{NOMA}} > R_{\text{TDMA}}$, i.e.,

$$\Big(1 + \rho\beta_{is}Y_i - T\Big(\frac{\beta_{is}}{\beta_{iw}} - 1\Big)\Big)^2 > \Big(1 + \beta_{is}Y_i\rho\Big)\Big(1 + \beta_{iw}Y_i\rho\Big). \tag{5.43}$$

Re-using the notation, $\beta_{iw}Y_i = \frac{T}{\rho} + \zeta$, the inequality in (5.43) becomes

$$\begin{aligned}&\Big(1 - \frac{T(\beta_{is} - \beta_{iw})}{\beta_{iw}} + \rho\beta_{is}\Big(\frac{T}{\rho\beta_{iw}} + \frac{\zeta}{\beta_{iw}}\Big)\Big)^2 > \\ &1 + \beta_{is}\rho\Big(\frac{T}{\rho\beta_{iw}} + \frac{\zeta}{\beta_{iw}}\Big) + \rho\Big(\frac{T}{\rho} + \zeta\Big) + \rho^2\Big(\frac{T}{\rho} + \zeta\Big)\beta_{is}\Big(\frac{T}{\rho\beta_{iw}} + \frac{\zeta}{\beta_{iw}}\Big).\end{aligned} \tag{5.44}$$

Multiplying through by β_{iw}^2 and after some manipulations, we obtain

$$\big(\beta_{is}\beta_{iw} - \beta_{iw}^2\big)T + \rho\zeta\big(\beta_{is}\beta_{iw} - \beta_{iw}^2\big) + \rho^2\zeta^2\big(\beta_{is}^2 - \beta_{is}\beta_{iw}\big) - T^2\big(\beta_{is}\beta_{iw} - \beta_{iw}^2\big) > 0. \tag{5.45}$$

After further manipulations, we obtain

$$\rho^2\beta_{is}\zeta^2 + \rho\beta_{iw}\zeta - \beta_{iw}T(T-1) > 0, \tag{5.46}$$

which is negative at $\zeta = 0$. Hence the cluster rate exceeds the TDMA rate at $\zeta > \zeta_0$, where ζ_0 is the larger root of (5.46), i.e., when

$$\Big(\beta_{iw}Y_i - \frac{T}{\rho}\Big) > \left(\frac{-\rho\beta_{iw} + \sqrt{\rho^2\beta_{iw}^2 + 4\beta_{iw}T(T-1)\rho^2\beta_{is}}}{2\rho^2\beta_{is}}\right). \tag{5.47}$$

Note that $\zeta_0 > 0$, so $\Big(\beta_{iw}Y_i - \frac{T}{\rho}\Big) > \zeta_0$ which yields $\beta_{iw}Y_i > \frac{T}{\rho}$. Thus the requirement

that $\beta_{iw}Y_i > \frac{T}{\rho}$ can be dropped in favor of

$$\beta_{iw}Y_i > \frac{T}{\rho} + \left(\frac{-\rho\beta_{iw} + \sqrt{\rho^2\beta_{iw}^2 + 4\beta_{iw}T(T-1)\rho^2\beta_{is}}}{2\rho^2\beta_{is}} \right). \tag{5.48}$$

After some manipulations, (5.48) becomes

$$Y_i > \frac{T}{\rho\beta_{iw}} + \left(\frac{-1 + \sqrt{1 + 4(\beta_{is}/\beta_{iw})T(T-1)}}{2\rho\beta_{is}} \right). \tag{5.49}$$

For notational simplicity, we define Y_0 equal to the right hand side of (5.49). Thus, if (5.49) holds, i.e., $Y_i > Y_0$, then a NOMA cluster will exceed the TDMA rate.

5.4.3 SNR/SINR distributions and sum rate

In this section, we derive the exact SNR/SINR distributions of singletons and strong and weak users in clusters in order to bound performance. Further, using the SNR/SINR distribution results we compute the expected sum rate of W-QoS NOMA in closed form. To achieve this, we assume perfectly correlated strong and weak user channels in a cluster.

From Sec. 5.2.1 and standard ZF results, for a singleton, $Y_i \sim \chi_p^2$, for $i = N_1 + 1, \ldots, N$. Denoting the CDF of a χ_p^2 variable as $F_p(\cdot)$, we have $P(Y_i \leq b) = F_p(b)$ for $i = N_1 + 1, \ldots, N$. Within a cluster, Y_i is constrained by $Y_i > Y_0$ and Y_i has the conditional CDF defined by $P(Y_i \leq b | Y_i > Y_0) = \frac{F_p(b) - F_p(Y_0)}{1 - F_p(Y_0)}$ for $b \geq Y_0$ and $i = 1, \ldots, N_1$. Using these results in (5.16), (5.37) and (5.39), we can write

$$P(\mathrm{SNR}_i < b) = F_p\Big(\frac{b}{\rho\beta_i}\Big), \;\; i = N_1 + 1, \ldots, N, \tag{5.50}$$

$$\mathrm{SINR}_{iw} = T, \;\; i = 1, \ldots, N_1, \tag{5.51}$$

$$P(\mathrm{SNR}_{is} < b) = \frac{F_p\left(\dfrac{b\beta_{iw}(T+1) + \beta_{is}T}{\rho\beta_{is}\beta_{iw}} \right) - F_p(Y_0)}{1 - F_p(Y_0)}, \tag{5.52}$$

$i = 1, \ldots, N_1$. It is worth pointing out that (5.50), (5.51) and (5.52) are the exact

closed form expressions for SNR/SINR distributions of singletons and weak and strong users of clusters, respectively.

Using the SNR/SINR distributions in (5.50)-(5.52) we are able to compute the expected sum rate of W-QoS NOMA in closed form. First, we write

$$E[R_{\text{sum}}] = \sum_{i=1}^{N_1} E[R_{c,i}] + \sum_{i=N_1+1}^{N} E[R_i], \tag{5.53}$$

where $E[R_i] = E[\log(1+\text{SNR}_i)]$ is the singleton rate, $E[R_{c,i}] = E[\log(1+\text{SNR}_{is})] + E[\log(1+\text{SNR}_{iw})]$ is the cluster rate and the SNR/SINRs are in (5.50)-(5.52). Now, the singleton rate can be obtained from [116, Eq. 7] which gives $E[\log(1+\rho Y)] = C(M,\rho)$ in closed form, when $Y \sim \chi^2_M$. $C(M_T,\rho)$ can be obtained from [116, Eq. 7] as

$$\begin{aligned} C(M_T,\rho) &= \frac{1}{\rho^{M_T}(M_T-1)!\,\ln 2}\int_0^{\infty} y^{(M_T-1)}e^{-y/\rho}\ln(1+y)dy \\ &= \frac{1}{\ln 2}\left(1+\sum_{i=1}^{M_T-1}\frac{1}{(M_T-i)!}\Big(-\frac{1}{\rho}\Big)^{M_T-i}\right)e^{\frac{1}{\rho}}E_1\Big(\frac{1}{\rho}\Big) \\ &+ \frac{1}{\ln 2}\sum_{i=1}^{M_T-1}\sum_{k=1}^{M_T-i}\sum_{l=1}^{k}\frac{(-1)^{M_T-i-k}\rho^{i+l-M_T}}{k(M_T-i-k)!(k-l)!}, \end{aligned} \tag{5.54}$$

where $E_1(x)$ is an exponential integral and $E_1(x) = \int_1^{\infty} e^{-xt}t^{-1}dt$. Using this result, the singleton rate in (5.17) can be written as

$$E[R_i] = E[\log(1+\rho\beta_i Y_i] = C(p,\rho\beta_i), \tag{5.55}$$

where $p = M_T - N + 1$.

The rate for a cluster depends on whether the scheduled cluster supports a pair of users or if the weak user was dropped leading to a singleton. If $Y_i > Y_0$ then a NOMA cluster is possible and the rate equals the sum of the strong and weak user rates. On the other hand, if $Y_i \leq Y_0$ then the cluster is dropped. Hence,

$$E[R_{c,i}] = P(Y_i > Y_0)E[R_{c,i}|Y_i > Y_0] + P(Y_i \leq Y_0)E[R_{c,i}|Y_i \leq Y_0]. \tag{5.56}$$

The first conditional mean rate in (5.56) can be written as

$$E[R_{c,i}|Y_i > Y_0] = E\Big[\log\Big(1 + \rho\beta_{is}Y_i - T(\frac{\beta_{is}}{\beta_{iw}} - 1)\Big)\Big|Y_i > Y_0\Big]. \tag{5.57}$$

Similarly, the second conditional mean rate can be written as

$$E[R_{c,i}|Y_i \leq Y_0] = E\Big[\log\Big(1 + \rho\beta_{is}Y_i\Big)\Big|Y_i \leq Y_0\Big]. \tag{5.58}$$

Equations (5.57) and (5.58) are derived in Appendix A giving the final result,

$$E[R_{c,i}] = I_1(a_i, b_i) + C(p, b_i) - I_1(1, b_i), \tag{5.59}$$

where $a_i = 1 - T\Big(\frac{\beta_{is}}{\beta_{iw}} - 1\Big)$, $b_i = \rho\beta_{is}$, $C(\cdot,\cdot)$ is [116, Eq. 7] and $I_1(\cdot,\cdot)$ is the closed form function given in (B.8).

Using (5.55) and (5.59) the expected sum rate becomes

$$\begin{aligned} E[R_{sum}] &= \sum_{i=N_1+1}^{N} C(p, b_i) + \sum_{i=1}^{N_1}\Big(I_1(a_i, b_i) + C(p, b_i) - I_1(1, b_i)\Big) \\ &= \sum_{i=1}^{N} C(p, b_i) + \sum_{i=1}^{N_1}\Big(I_1(a_i, b_i) - I_1(1, b_i)\Big). \end{aligned} \tag{5.60}$$

This appears to be the only analysis of a NOMA-ZFBF system which includes power allocation, guarantees a NOMA rate is higher than the TDMA rate and gives an exact upper bound for the sum rate.

5.4.4 Numerical results

In the numerical results the following parameter values and simulation settings are used. The system dimension controlled by L, M, N, N_1 and N_2 is defined in each figure. The noise power is set to $\sigma^2 = 1$ without loss of generality. For simplicity, we assume Rayleigh fading channels that have the same variance for both strong users and singletons. In particular, we take $\beta_i = \beta_{is} = 4$ and $\beta_{iw} = 1$ unless otherwise

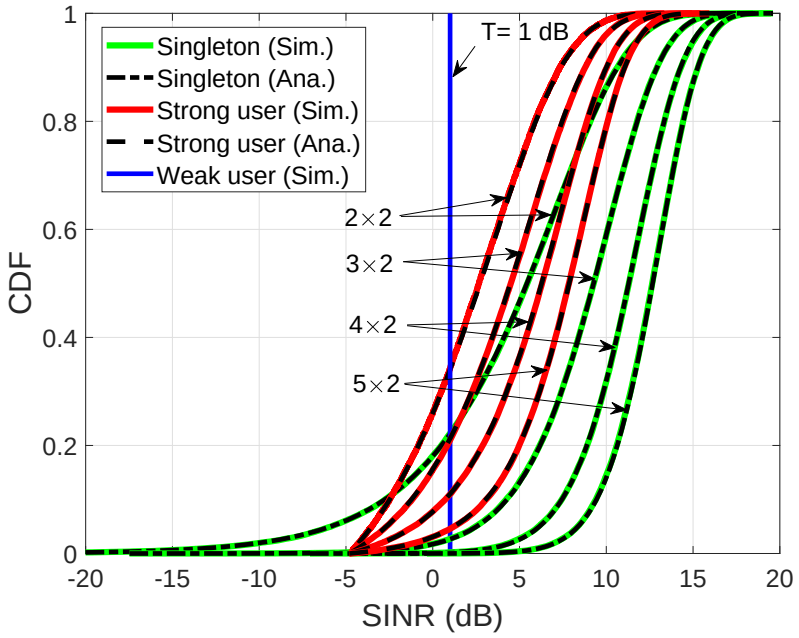


Figure 5-7: Analytical and simulated SNR/SINR CDFs of singletons and clustered strong users for various transmitting antennas M_T. $N = 2$ ($N_1 = 1$ & $N_2 = 1$), $\rho = 5$ dB and $T = 1$ dB.

specified in the figure legends. The values of SNR, ρ, weak user target, T, and the clustering parameters, Ψ_0 and Γ, are specified in each figure where relevant. The number of simulations for each CDF and for each point on a mean sum-rate curve is 100000.

In this section, we demonstrate the validity of the derived SNR/SINR distributions of singletons and clustered strong and weak users and finally the expected sum rate of the system. Moreover, the simulation validates the conclusions from the analysis: the significance of the correlation factor and channel power differences. Further, the analytical results below are based on the SINR results in Sec. 5.4 assuming perfectly correlated strong and weak users unless otherwise specified.

In Fig. 5-7, we observe that the analytical CDFs, (Ana.), of the output SNR of singletons, using (5.50), and clustered strong users, using (5.52), perfectly match the simulated CDFs, labelled (Sim.). Note that the SINR of weak users is set to equal the threshold value, T, as shown in Fig. 5-7. Further, we demonstrate the effect of the transmitting array size on the singleton and strong user SINRs, where the system

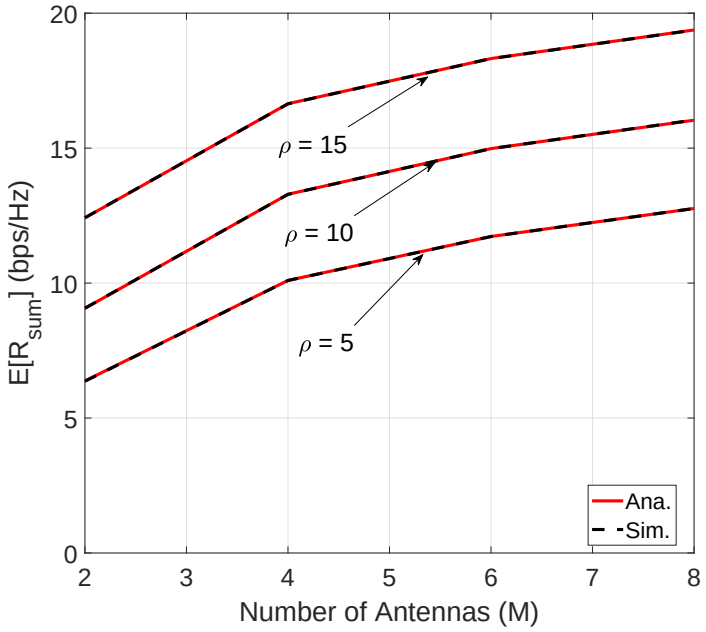


Figure 5-8: Analytical and simulated mean sum rates versus transmitting antennas M_T. $N = 2$ ($N_1 = 1$ & $N_2 = 1$) and $T = 1$.

dimension is denoted as $M_T \times N$. Similarly, in Fig. 5-8 we observe that the analytical mean sum rate (Ana.) matches the simulated mean sum rate (Sim.). It is obvious that as the number of transmitting antennas increases, the rates of strong users and singletons also increase. Note that the analysis above is independent of the number of users and will not have any impact on the results.

Fig. 5-9 presents rate CDFs of singletons, strong users, weak users and clusters with varying target rates for the weak user. The simulated rate CDFs of singletons, strong users and weak users employ equations (5.50), (5.52) and (5.51), respectively and the cluster rate is the sum of the strong and weak user rate. The strong user performance decreases with an increase in the target weak user rate. This is because an increase in target rate means that the power scaling factor for the weak user, $(1 - \alpha_i^2)$, is increased, thereby decreasing the power scaling factor of the strong user, α_i^2.

In Fig. 5-10, we compare the cluster rates from (5.42) with the equivalent TDMA rate. In addition, the cluster rate is computed using two different choices of power allocation techniques for α_i from [40] and [96]. These schemes keep the sum rate of a

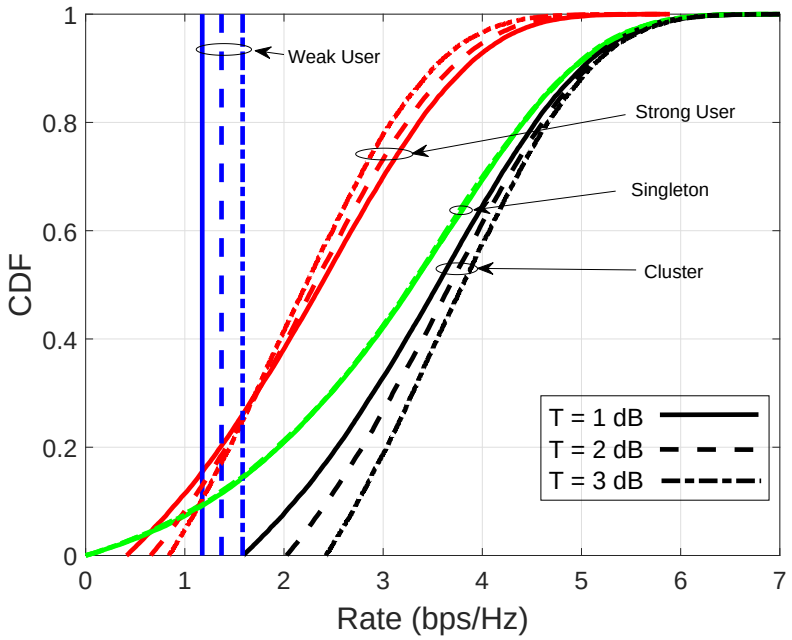


Figure 5-9: Rate CDFs of singleton, strong user, weak user and cluster. $M_T = 2$, $N = 2$ ($N_1 = 1$ & $N_2 = 1$) and $\rho = 5$ dB.

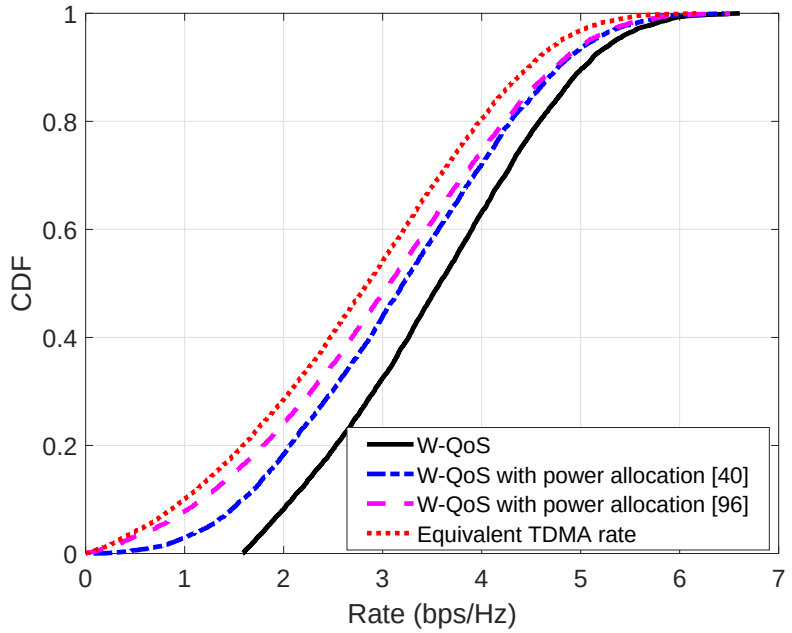


Figure 5-10: CDF plots for cluster-rate with W-QoS NOMA , it's equivalent TDMA rate and W-QoS NOMA with power allocation. $\rho = 5$ dB, $T = 1$ dB, $M_T = 2$ and $N = 2$ ($N_1 = 1$ & $N_2 = 1$).

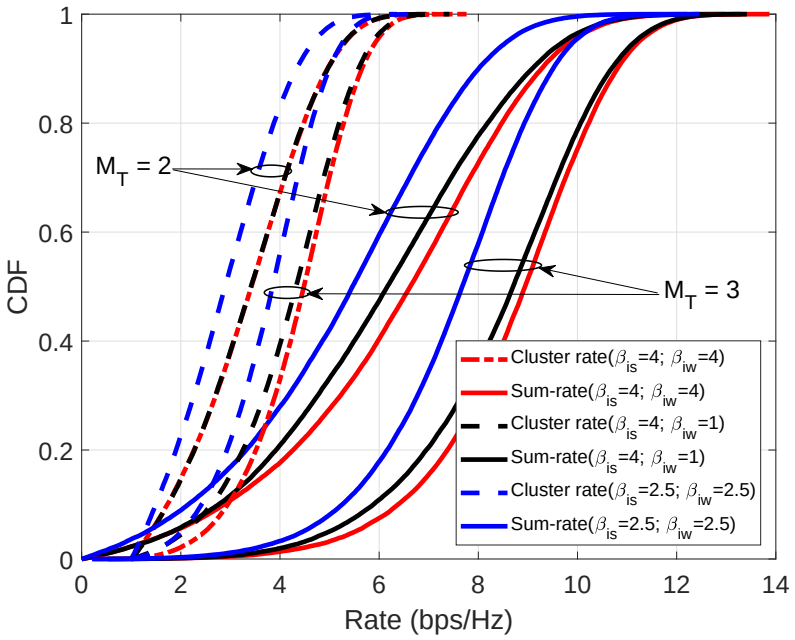


Figure 5-11: Rate CDFs. $N = 2$ ($N_1 = 1$ & $N_2 = 1$), $\rho = 5$ dB, and $T = 1$.

NOMA pair greater than that of the conventional multiuser BF system as described in section 5.2.4. The results are for $N = 2$ ($N_1 = 1$ and $N_2 = 1$) and the cluster rates are found from (5.23) for TDMA and (5.24) for NOMA with three different power allocation methods: W-QoS NOMA, [40] and [96]. The cluster rates of the W-QoS NOMA technique with all power allocation schemes are higher than the equivalent TDMA system, since this inequality is part of the design process. Also, the cluster rate with W-QoS NOMA is higher than the power allocation techniques from [40] and [96]. This follows from the design of W-QoS NOMA, where the power allocation is optimal under the given constraint. Moreover, we observe that cluster rate with power allocation according to [40] is better than that from [96]. From simulations we observe that [40] provides a larger fraction of the power to the weak user compared to [96].

In Fig. 5-11, we observe the rate dependence on the channel powers considering perfectly correlated strong and weak users in a cluster. Similar to the insights mentioned in Sec. 5.4, in Fig. 5-11 we observe that the cluster rate increases with an increase in the weak user channel power from $\beta_{iw} = 1$ to $\beta_{iw} = 4$. This is due to the

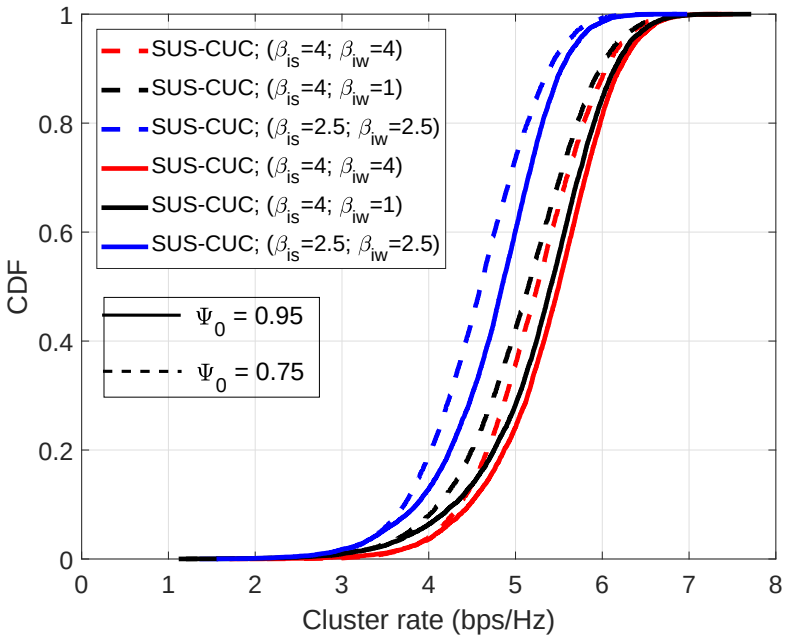


Figure 5-12: Cluster rate CDFs using SUS-CUC user selection. $M_T = 2$, $N = 2$, $L = 10$, $\rho = 5$ dB, $T = 1$ and $\Gamma = 0.4$. $N(N_1$ and $N_2)$ depend upon Ψ_0 and Γ.

fact that power allocation to the strong user increases with increase in weak user channel power, i.e, α_i^2 is an increasing function of β_{iw}. Further, in Fig. 5-12, to observe the importance of the correlation factor and the rate dependence on channel powers in practical clusters, we simulate W-QoS NOMA with SUS-CUC selected clusters varying the weak user channel powers, β_{iw}, and clustering correlation threshold, Ψ_0. We observe that the cluster rate increases with an increase in the correlation factor due to the decrease in intra-cluster interference. Moreover, we observe that power allocation to the strong user increases with an increase in the weak user channel power from $\beta_{iw} = 1$ to $\beta_{iw} = 4$. These observations are consistent with the analytical insights made in Sec. 5.4. Therefore, we can conclude that, if both channels of the users in a cluster are strong then this helps in clustering and overall cluster rate, and similarly, higher correlation helps to boost cluster rate.

To compare the analytical upper bound with the sum rate of W-QoS NOMA with different correlation thresholds, in Fig. 5-13, we plot the sum rate versus the transmit SNR, ρ. From this figure, we observe that the analysis gives a tight upper bound. Further, as shown in Sec. 5.4, higher correlation helps to boost the sum rate i.e., if

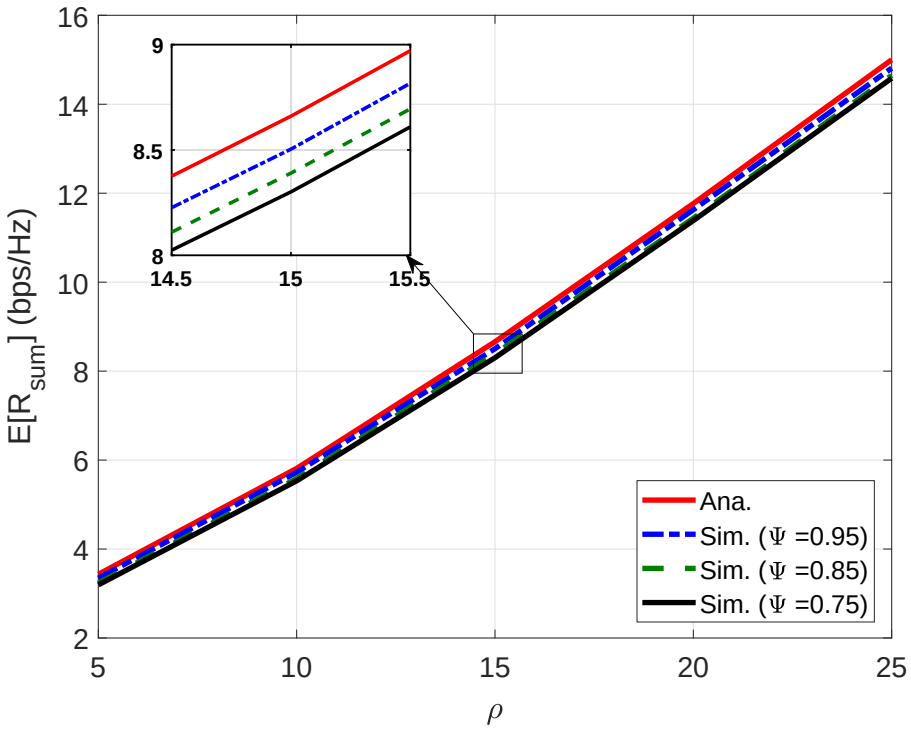


Figure 5-13: Sum rate of SUS-CUC user selection. $M_T = 2$, $T = 1$, $L = 10$ and $\Gamma = 0.4$. $N(N_1$ and $N_2)$ depend upon Ψ_0 and Γ.

we increase the correlation threshold from $\Psi_0 = 0.75$ to $\Psi_0 = 0.95$ then the sum-rate of SUS-CUC algorithm comes closer to analytical upper bound.

5.5 Conclusion

In NOMA, the main challenge is to pair users and allocate powers that can maximize the sum rate of the system. We proposed a simple power allocation technique that guarantees QoS to the weak user. Our proposed W-QoS NOMA algorithm was verified using two clustering techniques, SUS-CUC and SUS-UM. Compared to existing SUS-UM techniques in [96], the proposed SUS-UM integrated W-QoS NOMA approach achieves performance gains over SUS-UM [96] depending upon the number of participating users, transmit SNR and allocated weak user's rate. Further, the W-QoS NOMA method with modified SUS-CUC has a marginal performance gain over W-QoS NOMA based SUS-UM, due to its clustering technique.

Assuming an arbitrary correlation between users in a cluster, we have derived the SNR/SINR expressions of strong and weak users. These expressions show the importance of correlation factor which has a higher impact on the power penalty of the strong user, which in turn affects the cluster rate. Unlike other techniques, this analysis leads to the insight that a power differential is not required among the strong and weak user. In fact, both channels being strong helps clustering and the overall cluster rate. Further, assuming perfect correlation between strong and weak users in a cluster, this chapter has derived the exact SNR/SINR distributions of singletons and clustered strong and weak users. The analysis is exact for singletons and clusters with perfectly correlated channels. Finally, we have derived the closed form sum-rate of the system which includes power allocation, guarantees a NOMA rate higher than the TDMA rate and gives an exact upper bound for SNRs/SINRs and rate.

Chapter 6

On The Impact of Channels with Rician Fading and Spatial Correlation on NOMA Performance

6.1 Introduction

In future generation cellular communication with dense networks, different sets of propagation directions are likely and these propagation directions cause variations in the covariance profiles across different users. Furthermore, as discussed in chapter 5, NOMA-ZFBF has been identified as an enabling technology for 5G cellular communication. In multi-user NOMA-ZFBF systems two kinds of user groups are required: (a) those with good spatial separation (semi-orthogonal users (SU)) and (b) those with high correlation (highly correlated (HC) users) to achieve spectral efficiency [40]. In particular, NOMA groups correlated users into clusters and their information is transmitted in a single beam multiplexed in the power domain. ZF is used for the semi-orthogonal users [33] which eliminates the inter-cluster interference [40] [96].

Most of the recent works have shown that NOMA can achieve higher sum-rate and spectral efficiency compared to OMA systems [106, and references therein] [117]. The majority of the works have considered either I.I.D. Rayleigh fading [114]-[107] or

3GPP LTE channel models [118]-[119]. However, none of the above-mentioned works consider the effects of LoS components and spatial correlation while clustering, which are likely to be the dominant features in forthcoming millimeter wave and massive MIMO based cellular communication systems. The exception is [120] which has considered Rician fading channels in NOMA-based cooperative relaying systems. To the best of our knowledge, no work has been published considering these components for different channel models. Such a comparison will give insights into the effect of LoS channels and spatial correlations on NOMA systems.

Therefore, in this chapter, we identify a propagation scenario which is likely to provide both similar and near-orthogonal channel vectors, i.e., a scenario which is good for NOMA performance. In order to achieve this, first, we consider a propagation scenario suitable for dense networks, which is modeled via the semi-correlated Rician distribution, where each user has a different covariance matrix. Then we explore and analyze the variance of the interference power between two users, where a large variance indicates this channel behavior is more likely to give good NOMA performance. Three key properties of the analytical variance are identified and discussed for the uncorrelated channels and correlated channels with equal and unequal correlations. Second, we explore the effects of Rician (K) factor, central azimuth angle and angular spread (AS) which causes variations in the scattering, leading to the variation in the spatial correlation of users. Furthermore, the Rician and Rayleigh channels in correlated and I.I.D. environments are compared to provide some new insights into the role of LoS and correlated components with respect to NOMA beam formation and clustering of HC users. Finally, we explore the effect of channel models on NOMA-cluster formation and cluster rate. For this, we have chosen the user selection and power scheduling technique proposed in [96]. This approach adopted the semiorthogonal user selection (SUS) algorithm from [33] and adopted its own user matching (UM) technique, for convenience referred to here as the SUS-UM algorithm. Numerical results show the impact of LoS and spatial correlation among the users. We demonstrate that correlated Rayleigh and Rician channels with similar channel statistics achieve higher variance which in turn enhances the performance of

the system.

6.2 Downlink NOMA system model

We consider the downlink of a single base station (BS), multiuser ZFBF system as proposed in [96]. The BS is equipped with M_T antennas and Ω users are selected for processing from L possible candidates. We assume that all users are equipped with a single antenna. To integrate NOMA with a multiuser system scenario, the Ω users are grouped into N_1 clusters with $B = 2$ highly correlated users in each cluster and N_2 singletons, i.e., $\Omega = N_1 B + N_2$, using the SUS-UM technique. Here the singletons are the users that can't be grouped into a correlated cluster. As a single beam is used to transmit information to a cluster, we have $N = N_1 + N_2$ beams, where $M_T \geq N$. The $N \times M_T$ channel matrix from the BS to the strong users in N_1 clusters and N_2 singletons is given by

$$\mathbf{H} = [\underbrace{\mathbf{h}_{1s}^{\mathrm{T}}, \mathbf{h}_{2s}^{\mathrm{T}}, \cdots, \mathbf{h}_{N_1 s}^{\mathrm{T}}}_{\text{Clusters}}, \underbrace{\mathbf{h}_{N_1+1}^{\mathrm{T}}, \cdots, \mathbf{h}_{N}^{\mathrm{T}}}_{\text{Singletons}}]^{\mathrm{T}}. \tag{6.1}$$

We use subscripts s and w to denote the strong/near and weak/far users in the N_1 clusters, respectively. Moreover, $\mathbf{h}_{is}$ represents the channel gain vector of the strong user in the i^{th} cluster and $\mathbf{h}_i$ is the channel vector of the i^{th} singleton.

We assume spatially correlated Rician fading channels with unequal statistics for each user. Hence, a generic channel can be written as

$$\mathbf{h} = \sqrt{\beta}\mathbf{g}, \tag{6.2}$$

where β is the link gain and $\mathbf{g}$ is the normalized channel denoted by

$$\mathbf{g} = \sqrt{\frac{K}{K+1}}\ \bar{\mathbf{h}} + \sqrt{\frac{1}{K+1}}\ \hat{\mathbf{h}}\mathbf{R}^{1/2}, \tag{6.3}$$

where $\bar{\mathbf{h}}$, $\hat{\mathbf{h}}$ and $\mathbf{R}$ are the specular (LOS), diffuse (scattered) components of the channel and the $M_T \times M_T$ transmit spatial correlation matrix, respectively. The ratio

between the powers of the specular and scattered components is defined as the Rician factor (K). Note that K and $\mathbf{R}$ are unique to each user. Assuming a uniform linear array (ULA), the LOS component is $\bar{\mathbf{h}} = [1, e^{j2\pi d \sin\bar{\phi}}, \cdots, e^{j2\pi(M_T-1)d\sin\bar{\phi}}]$, where $\bar{\phi}$ is the angle-of-departure (AoD) of the specular component and d is the inter-element spacing in wavelengths.

We model the distribution of users with a uniform density over the coverage area of the cell as in [104]. The link gains are $\beta = A\xi(r_0/r)^{-\gamma}$, where A is the unit-less constant for the geometric attenuation at a reference distance of r_0, r is the distance between the BS and the user, γ is the attenuation exponent and $\xi = 10^{(X/10)}$, where X depends upon the shadow fading standard deviation (σ_{sf}^2), i.e., $X \sim \mathcal{N}(0, \sigma_{sf}^2)$.

The transmitter performs ZFBF based on $\mathbf{H}$. The un-normalized precoding matrix, $\mathbf{W}_0$, is given by

$$\mathbf{W}_0 = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1} = [\mathbf{w}_{01}, \mathbf{w}_{02}, \cdots, \mathbf{w}_{0N}], \tag{6.4}$$

so that the normalized precoder is $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N]$, where $\mathbf{w}_i = \mathbf{w}_{0i}/||\mathbf{w}_{0i}||$. Now, the received signals can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{W}\mathbf{x} + \mathbf{z}, \tag{6.5}$$

where $\mathbf{x}$ is the signal to be transmitted and $\mathbf{z}$ denotes the noise vector. Further, $\mathbf{x}$ can be represented as $\mathbf{x} = [x_1, \cdots, x_{N_1}, x_{N_1+1}, \cdots, x_N]^{\mathrm{T}}$, where $x_1, \cdots, x_{N_1}$ are the signals for clusters given by $x_i = \left(x_{is}\alpha_i + x_{iw}\sqrt{1-\alpha_i^2}\right)$ and $x_{N_1+1}, \cdots, x_N$ are the signals for singletons. Further, x_{is} and x_{iw} are the strong user signal and weak user signal, respectively. α_i and $\sqrt{1-\alpha_i^2}$ are the amplitude scalings and α_i^2 and $1-\alpha_i^2$ are the power fractions of x_{is} and x_{iw}, respectively. The signal strength is $E_x = E[|x_i|^2]$ and the transmit SNR is $\rho = \frac{E_x}{\sigma^2}$, where σ^2 is noise variance. We assume that $\sigma^2 = 1$. The SNR of the strong user in a cluster has no inter-cluster interference as the ZF precoders are designed based on the strong users' channels. Further, successive interference cancellation (SIC) is used to first detect the weak user's signal. After successful SIC of the weak user signal, the SNR of the strong user

is given by

$$\mathrm{SNR}_{is} = |\mathbf{h}_{is}\mathbf{w}_i|^2\alpha_i^2\rho, \quad i = 1, 2, \cdots, N_1. \tag{6.6}$$

Since ZF is only performed with the strong users' channels, the weak user suffers from interference from other clusters and from their strong user partner. Hence, the SINR of the weak user is given by

$$\mathrm{SINR}_{iw} = \frac{|\mathbf{h}_{iw}\mathbf{w}_i|^2(1-\alpha_i^2)\rho}{|\mathbf{h}_{iw}\mathbf{w}_i|^2\alpha_i^2\rho + \sum\limits_{j\neq i}|\mathbf{h}_{iw}\mathbf{w}_j|^2\rho + 1}, \quad i = 1, 2, \cdots, N_1. \tag{6.7}$$

As such, the downlink spectral efficiency for the i-th cluster is given by $\mathrm{R}_{ci} = \log_2(1+\mathrm{SNR}_{is}) + \log_2(1+\mathrm{SINR}_{iw})$, where $i = 1, 2, \cdots, N_1$.

6.3 Variance analysis

Let $\mathbf{h}_i$ and $\mathbf{h}_j$ be the channels of two generic users before the SUS selection procedure classifies them as selected, non-selected, strong, weak, etc. Define $P_{ij} = |\mathbf{h}_i\mathbf{h}_j^H|^2$ as the power of $\mathbf{h}_j$ in the direction of $\mathbf{h}_i$. If $P_{ij} = 0$, both channels are orthogonal to each other and separate beams are best for these users. On the other hand, if P_{ij} is large the channels are correlated and both users can be supported in a single beam via NOMA. Hence, if $\mathrm{Var}(P_{ij})$ is large, it is more likely to achieve both SU and HC users which assist NOMA. Therefore, $\mathrm{Var}(P_{ij})$ is analyzed here for the Rician channel model.

The Rician channel model in (6.3) can be written as

$$\mathbf{g}_i = \mathbf{a}_i + b_i\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}. \tag{6.8}$$

Considering $\beta = 1$ and expanding $|\mathbf{h}_i\mathbf{h}_j^H|^2$ gives 16 terms but these collapse to

$$E[P_{ij}] = |\mathbf{a}_i\mathbf{a}_j^H|^2 + b_j^2\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H + b_i^2\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H + b_i^2b_j^2\,\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j), \tag{6.9}$$

using the known results [121]

$$E[\ \hat{\mathbf{h}}_i] = \mathbf{0}, \tag{6.10a}$$

$$E[\ \hat{\mathbf{h}}_i^H\ \hat{\mathbf{h}}_i] = \mathbf{I}, \tag{6.10b}$$

and

$$E[\ \hat{\mathbf{h}}_i\mathbf{A}\ \hat{\mathbf{h}}_i^H] = \mathrm{tr}(\mathbf{A}). \tag{6.10c}$$

The variance can be written as

$$\mathrm{Var}[P_{ij}] = E[P_{ij}^2] - E[P_{ij}]^2. \tag{6.11}$$

The second term in (6.11) is given in (6.9). Hence, we need $E[P_{ij}^2]$ to complete the variance calculation. Now P_{ij}^2 contains 256 terms so a direct expansion is lengthy. Instead, denote $\mathbf{Q} = \mathbf{h}_j^H\mathbf{h}_j$ and we can write

$$E[P_{ij}^2] = E\Big[E[\mathbf{h}_i\mathbf{Q}\mathbf{h}_i^H\mathbf{h}_i\mathbf{Q}\mathbf{h}_i^H]\Big], \tag{6.12}$$

where the inner expectation is over $\mathbf{h}_i$ and the outer expectation is over $\mathbf{h}_j$. This breaks up the calculations into manageable steps. Using the results

$$E[\ \hat{\mathbf{h}}_i^H\ \hat{\mathbf{h}}_i\mathbf{A}\ \hat{\mathbf{h}}_i^H\ \hat{\mathbf{h}}_i] = \mathbf{A} + \mathrm{tr}(\mathbf{A})\mathbf{I}, \tag{6.13a}$$

and

$$E[(\ \hat{\mathbf{h}}_i\mathbf{A}\ \hat{\mathbf{h}}_i^H)^2] = \mathrm{tr}(\mathbf{A}^2) + (\mathrm{tr}(\mathbf{A}))^2, \tag{6.13b}$$

in addition to (6.10) gives

$$\begin{aligned} E[P_{ij}^2] = E\Big[&(\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H)^2 + 2b_i^2\mathbf{a}_i\mathbf{Q}\mathbf{a}_i\mathrm{tr}(\mathbf{Q}\mathbf{R}_i) + \\ &2b_i^2\mathbf{a}_i\mathbf{Q}\mathbf{R}_i\mathbf{Q}\mathbf{a}_i^H + b_i^4\mathrm{tr}(\mathbf{Q}\mathbf{R}_i\mathbf{Q}\mathbf{R}_i) + b_i^4\Big(\mathrm{tr}(\mathbf{Q}\mathbf{R}_i)\Big)^2\Big]. \end{aligned} \tag{6.14}$$

In order to solve (6.14) we derive three results. The first is

$$\begin{aligned} E[\mathbf{QBQ}] &= \mathbf{a}_j^H\mathbf{a}_j\mathbf{B}\mathbf{a}_j^H\mathbf{a}_j + b_j^2\mathbf{a}_j^H\mathbf{a}_j\mathbf{B}\mathbf{R}_j + b_j^2\mathrm{tr}(\mathbf{B}\mathbf{R}_j)\mathbf{a}_j^H\mathbf{a}_j \\ &+ b_j^2\mathbf{a}_j\mathbf{B}\mathbf{a}_j^H\mathbf{R}_j + b_j^2\mathbf{R}_j\mathbf{B}\mathbf{a}_j^H\mathbf{a}_j + b_j^4\mathbf{R}_j\mathbf{B}\mathbf{R}_j + b_j^4\mathrm{tr}(\mathbf{B}\mathbf{R}_j)\mathbf{R}_j. \end{aligned} \tag{6.15}$$

Derivation of (6.15) simply expands **QBQ** and uses the results in (6.10) and (6.13). Secondly,

$$E[\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H\mathrm{tr}(\mathbf{Q}\mathbf{R}_i)] = \mathbf{a}_iE[\mathbf{h}_j^H\mathbf{h}_j\mathbf{R}_i\mathbf{h}_j^H\mathbf{h}_j]\mathbf{a}_i^H. \tag{6.16}$$

Thirdly,

$$E[(\mathrm{tr}(\mathbf{Q}\mathbf{R}_i))^2] = \mathrm{tr}\Big(\mathbf{R}_iE[\mathbf{h}_j^H\mathbf{h}_j\mathbf{R}_i\mathbf{h}_j^H\mathbf{h}_j]\Big). \tag{6.17}$$

Note that (6.16) and (6.17) have expectations of the form E[**QBQ**] so are known from (6.15). Using (6.15), (6.16) and (6.17) gives[1]

$$\begin{aligned} \mathrm{Var}[P_{ij}] &= 2b_j^2|\mathbf{a}_i\mathbf{a}_j^H|^2\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H + b_i^4(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)^2 \\ &+ b_j^4(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H)^2 + 2b_i^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)|\mathbf{a}_i\mathbf{a}_j^H|^2 \\ &+ 8b_i^2b_j^2Re\Big\{(\mathbf{a}_i\mathbf{a}_j^H)(\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j\mathbf{a}_i^H)\Big\} \\ &+ 2b_i^2b_j^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H) + 4b_i^2b_j^4\mathbf{a}_i\mathbf{R}_j\mathbf{R}_i\mathbf{R}_j\mathbf{a}_i^H \\ &+ 4b_i^4b_j^2\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j\mathbf{R}_i\mathbf{a}_j^H + 2b_i^2b_j|\mathbf{a}_i\mathbf{a}_j^H|^2\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) \\ &+ 2b_i^2b_j^4(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H)\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) + 2b_i^4b_j^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) \\ &+ 2b_i^4b_j^4\mathrm{tr}[(\mathbf{R}_i\mathbf{R}_j)^2] + b_i^4b_j^4\mathrm{tr}[(\mathbf{R}_i\mathbf{R}_j)]^2. \end{aligned} \tag{6.18}$$

Equal and unequal correlated Rayleigh

Note that, equal correlated Rayleigh means equal level of transmit spatial correlation at both users and unequal correlated Rayleigh means each user has unequal level of transmit spatial correlation structures. If we consider pure NLoS conditions (i.e., $K_i = K_j = 0$) then (6.18) reduces to the unequal correlated Rayleigh case and

$$\mathrm{Var}[P_{ij}] = 2 \times \mathrm{tr}[(\mathbf{R}_i\mathbf{R}_j)^2] + \mathrm{tr}[(\mathbf{R}_i\mathbf{R}_j)]^2. \tag{6.19}$$

[1]The detailed derivation of the variance is given in appendix C.1.

In (6.19), if we consider equal correlation matrices (i.e., $\mathbf{R}_i = \mathbf{R}_j = \mathbf{R}$) then (6.18) reduces to

$$\text{Var}[P_{ij}] = 2 \times \text{tr}[(\mathbf{R})^4] + \text{tr}[(\mathbf{R})^2]^2. \tag{6.20}$$

6.3.1 Discussion

Inspection of (6.18), (6.19) and (6.20) reveals three interesting properties.

Remark 1: Correlation increases the variance. This is easily seen in (6.20) where the $\mathbf{R} = \mathbf{I}$ of uncorrelated channels gives $\text{Var}[P_{ij}] = 2M_T + M_T^2$, while the $\mathbf{R} = 1_{M_T \times M_T}$ for perfectly correlated channels gives $\text{Var}[P_{ij}] = 3M_T^4$.

Remark 2: Equal correlation matrices increase the variance. In (6.18) and (6.20), $\text{tr}(\mathbf{R}_i \mathbf{R}_j)$ terms are found and it was demonstrated in [121] that these terms are maximized when $\mathbf{R}_i = \mathbf{R}_j$.

Remark 3: Similar LoS directions increase the variance. From (6.18) we see that the $|\mathbf{a}_i \mathbf{a}_j^H|$ terms are maximized by $\mathbf{a}_i = \mathbf{a}_j$. Further, it is likely that the terms involving $\mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H$ are also increased when $\mathbf{a}_i = \mathbf{a}_j$. This comment arises from the fact that the structure of $\mathbf{R}_i$ in some sense is based on the LoS direction (see (2.14) for example). Hence, similar LoS directions are likely to increase the variance.

In terms of NOMA systems, we conclude that correlated channels and similar channel statistics will enhance the performance.

6.4 Simulation results

In simulations, the user distances are randomly generated with a maximum distance of $r = 100$m with a reference distance $r_0 = 1$m. The other pathloss parameters are $\sigma_{sf} = 8$, and $\gamma = 3.5$, while the parameter A is chosen such that tenth percentile of the instantaneous SNR(dB) of ZFBF with four users selected by SUS [33] is 0 dB, when $M_T = 4$ and $\rho = 15$ dB. Further, we employ the one ring model for the spatial correlation matrices defined in Sec. 2.1.3 given by (2.14). We set $d = 0.25$ and assume both fixed and user-specific ϕ values. For the Rician case a fixed K-factor of 5 dB for each user is used unless otherwise specified. The LoS angle ($\bar{\phi}$) and the central

Table 6.1: $\widehat{\text{Var}}[P_{ij}]$ results with and without correlation.

I.I.D. Rayleigh	Corr. Rayleigh	I.I.D. Rician	Corr. Rician
24.78	245.68	27.27	54.16

Table 6.2: $\widehat{\text{Var}}[P_{ij}]$ results for different ϕ distributions.

	$\phi \sim \mathcal{U}[-\pi, \pi]$	$\phi \sim \mathcal{U}[-\pi/12, \pi/12]$	$\phi = 0$
Corr. Rayleigh	245.68	565.17	701.99
I.I.D. Rician	27.27	33.51	36.06

azimuth angle (ϕ), are assumed equal.

First, we verify the properties 1-3 discussed in Sec. 6.3. Note that $\text{Var}[P_{ij}]$ is given in (6.18) for fixed ϕ_i, Δ_i, ϕ_j, Δ_j, whereas ϕ and Δ vary spatially as well as over users. Hence, we compute $\text{Var}[P_{ij}]$ averaged over a variety of distributions for ϕ and Δ. If n samples of ϕ_i, Δ_i, ϕ_j, Δ_j, are generated giving $P_{ij}^{(i)}, \cdots, P_{ij}^{(n)}$ then using the conditional variance formula, we have the estimated variance:

$$\widehat{\text{Var}}[P_{ij}] = \sum_{l=1}^{\text{n}} \frac{\text{Var}[P_{ij}^{(l)}] + E[P_{ij}^{(l)}]^2}{\text{n}} - \left(\sum_{l=1}^{\text{n}} \frac{E[P_{ij}^{(l)}]}{\text{n}} \right)^2 \tag{6.21}$$

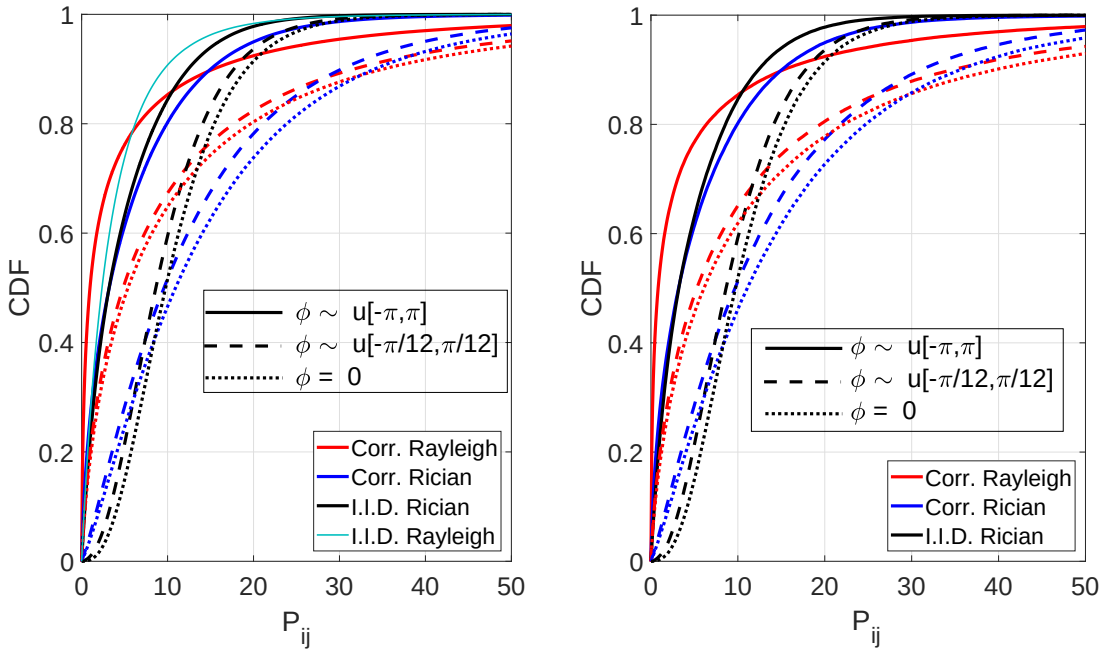
where (6.18) and (6.9) are used to compute $\text{Var}[P_{ij}]$ and $E[P_{ij}]$ and simulation averages over the varying correlation parameters.

Remark 1: Using (6.21) with $\text{n} = 10^6$ samples of $\phi \sim \mathcal{U}(-\pi, \pi)$, $\Delta \sim \mathcal{U}[-\pi/12, \pi/12]$ gives Table 6.1, where the increase in variance with correlation is very clear.

Remark 2: Using (6.21) with $\text{n} = 10^6$ samples of $\Delta \sim \mathcal{U}[-\pi/12, \pi/12]$ and varying ϕ gives Table 6.2. Row 1 of Table 6.2 shows the increase in variance as the correlation matrices become more similar with $\phi = 0$ giving the same central azimuth angle for both users.

Remark 3: The second row of Table 6.2 shows the increase in variance as the LoS angle become more aligned with $\phi = 0$ corresponding to perfect alignment.

Next, we take a more detailed look at the impact of LoS and correlation on P_{ij}. Fig. 6-1 illustrates the cumulative distribution function (CDF) of P_{ij} with different Δ and ϕ distributions. We observe that both correlated Rayleigh and correlated Rician channels with fixed central azimuth have a higher variance compared to their

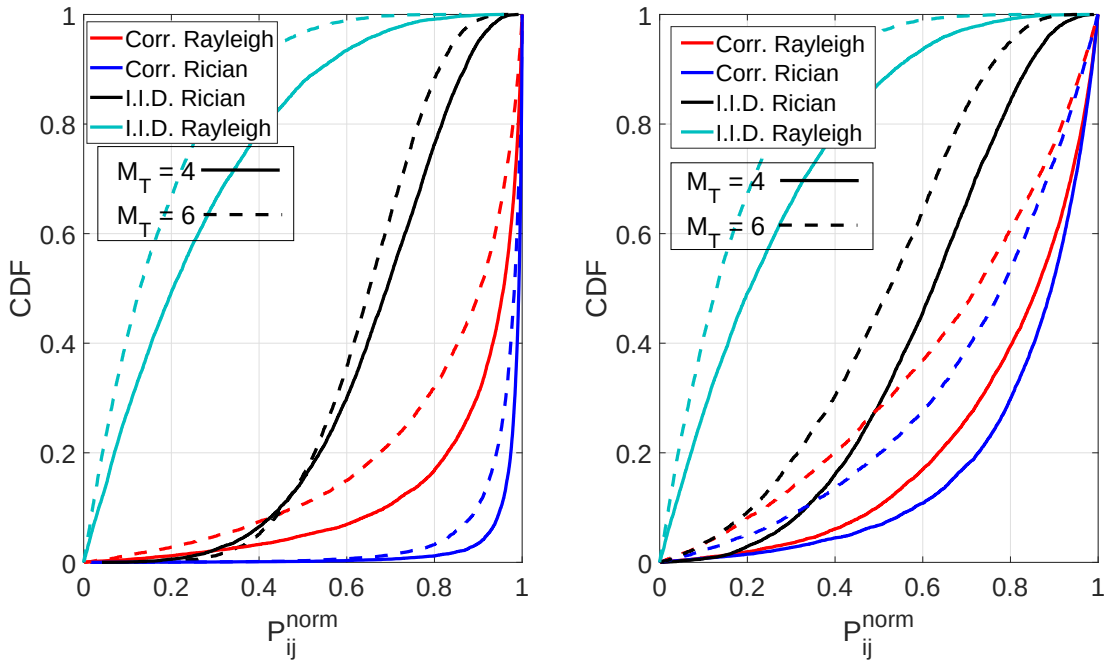


(a) $\Delta \sim \mathcal{U}[-\pi/6, \pi/6]$ (b) $\Delta \sim \mathcal{U}[-\pi/12, \pi/12]$

Figure 6-1: CDF of P_{ij} with $K = 5$ dB and $M_T = 4$.

respective I.I.D. channels. Similarly, transitioning from larger to smaller angular spread $\big(\Delta = \mathcal{U}(-\pi/6, \pi/6)$ to $\mathcal{U}(-\pi/12, \pi/12)\big)$ increases the variance. This is due to an increase in correlation of the channels. Also, as the range of ϕ decreases from 2π to $\pi/6$ to zero, the correlation matrices become more similar and variance increases. Finally, the variance of I.I.D. Rician channels increases when the LoS angles become more aligned, i.e., $\phi = 0$. These comments are all consistent with the properties 1-3. The CDFs clearly demonstrate that correlated Rayleigh is well-matched to NOMA scheduling. The CDFs rise rapidly giving many small values of P_{ij} (needed for beam construction) and then approach 1 much more slowly giving many high values of P_{ij} (needed for cluster formation).

In Fig. 6-2, we plot the CDFs of the normalized powers, $P_{ij}^{\text{norm}} = \big(|\mathbf{h}_i \mathbf{h}_j^H|/||\mathbf{h}_i||||\mathbf{h}_j||\big)^2$, in order to make a fair comparison between different system sizes, $M_T \in \{4, 6\}$. It is clear that the similarity of the channels decreases with an increase in M_T. Further, the correlated Rician channel is ideal for cluster formation but not for ZF beams as the normalized variance is predominantly close to 1. In contrast, the correlated



(a) $\phi = 0$ (b) $\phi \sim \mathcal{U}[-\pi/12, \pi/12]$

Figure 6-2: CDFs of P_{ij}^{norm} with $K = 5$ dB and $\Delta \sim \mathcal{U}[-\pi/12, \pi/12]$.

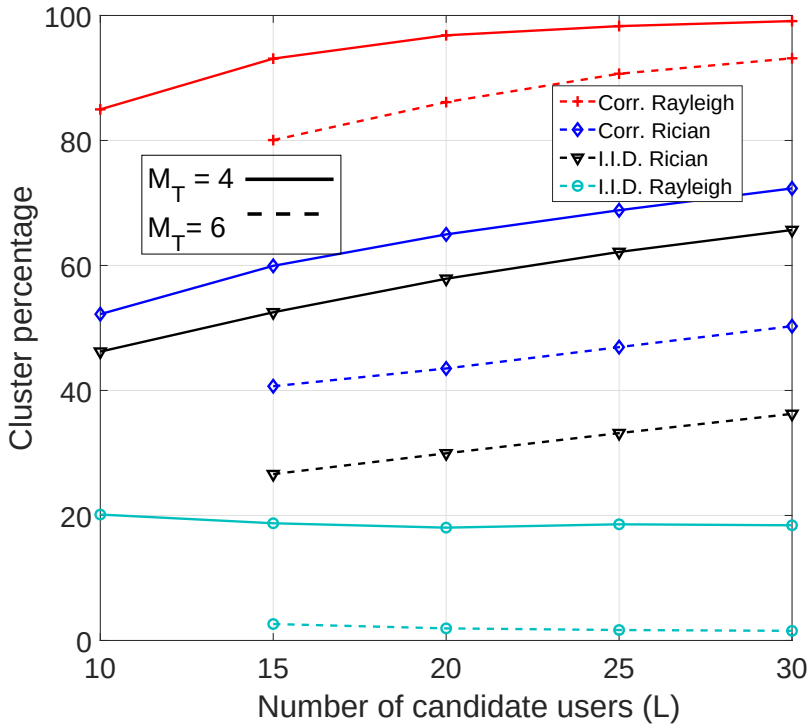


Figure 6-3: Cluster percentage achieved by SUS-UM vs the number of candidate users. $\Gamma = 0.4$, $K = 5$ dB, $\phi \sim \mathcal{U}[-\pi, \pi]$ and $\Delta \sim \mathcal{U}[-\pi/12, \pi/12]$.

Rayleigh channel is suitable to provide both SU and HC users. As in Fig. 6-1 we observe more similar channels for $\phi = 0$ than for $\phi = \mathcal{U}(-\pi/12, \pi/12)$ due to more similar correlation matrices and more aligned LoS vectors.

In Figs. 6-3 - 6-5, we investigate the impact of different channels on the particular NOMA technique labelled SUS-UM [96]. SUS-UM generates ZF beams using the SUS algorithm [33] followed by user-matching (UM) to create clustered users based on rate maximization. Note that the SUS technique accepts separate beams when the user channels have a correlation metric less than a threshold value, Γ.

Fig. 6-3 shows the cluster percentage achieved by the SUS-UM technique [96], i.e., the percentage of the separate beams which support clusters. We have discarded the cases when only 1 SU user is selected by the SUS algorithm to focus on genuine multi-user scenarios. Note that the similarity of the channels decreases with an increase in M_T and so the number of clusters. As can be seen, the cluster percentage decreases with an increase in M_T, this is similar to the observation made in Fig. 6-2. As expected, the highest percentage of clusters are formed with correlated Rayleigh channels due to its higher variance.

Fig. 6-4 illustrates the correlation between the strong and weak users within a cluster selected via SUS-UM. This intra-cluster correlation is given by

$$\big((|\mathbf{h}_{is}\mathbf{h}_{iw}^{H}|)/(||\mathbf{h}_{is}||||\mathbf{h}_{iw}||)\big). \tag{6.22}$$

As expected, we observe that intra-cluster correlation increases with an increase in the number of candidate users. Further, transitioning from larger to smaller angular spread increases the intra-cluster correlation. The correlated Rayleigh channel gives higher intra-cluster correlation due to its higher variance. Finally, Fig. 6-5 gives the mean cluster rate for all channel models according to the number of clusters formed. The correlated channels have better cluster-rate compared to their respective I.I.D. channel models and correlated Rayleigh has better cluster-rate compared to all other channels.

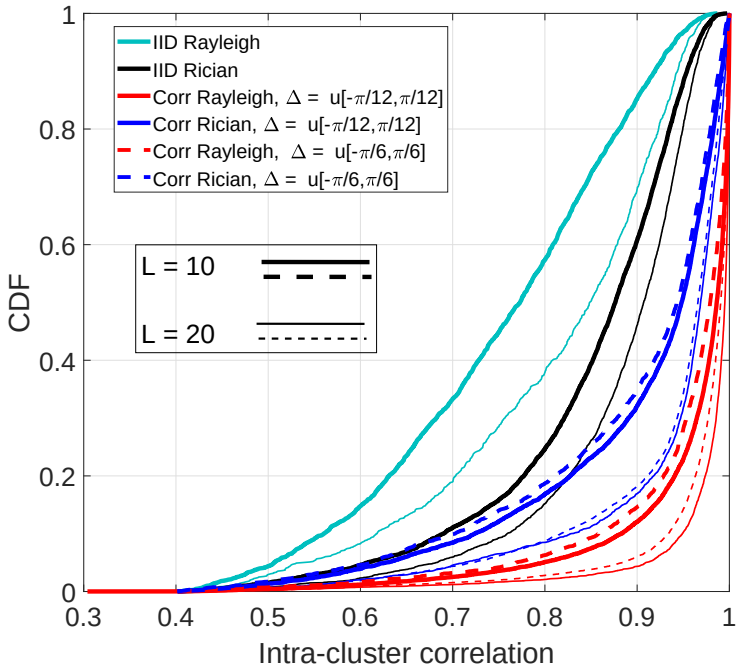


Figure 6-4: Intra-cluster correlation CDF for SUS-UM selected clusters. $\Gamma = 0.4$, $K = 5$ dB, $\phi \sim \mathcal{U}[-\pi, \pi]$.

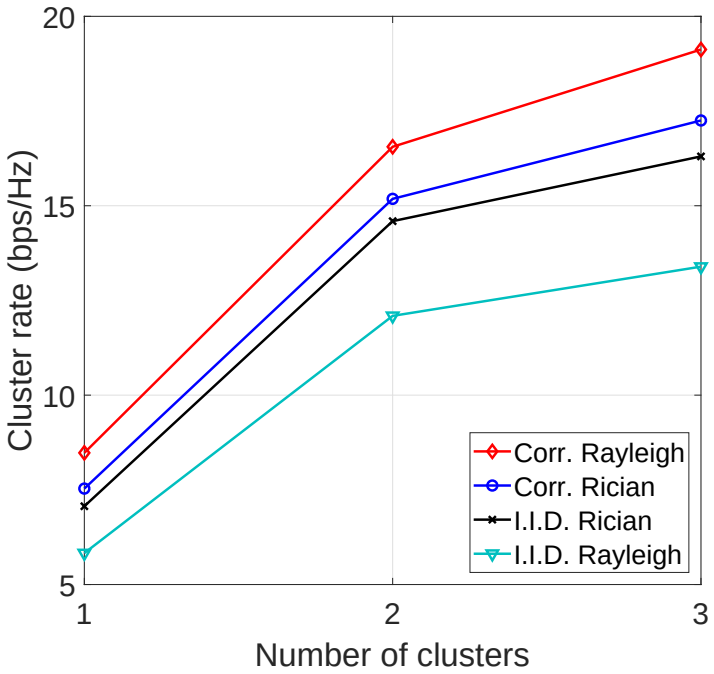


Figure 6-5: Mean cluster rate vs the number of clusters using SUS-UM. $\Gamma = 0.4$, $K = 5$ dB, $\phi \sim \mathcal{U}[-\pi, \pi]$, $\Delta \sim \mathcal{U}[-\pi/12, \pi/12]$, $\rho = 15$ dB, $M_T = 4$, and $L = 10$.

6.5 Conclusion

In this chapter, we have derived the variance of the interference power between two user channels for the correlated Rician fading channel. If the variance is large, the channel is more likely to produce both semi-orthogonal and highly correlated users which assist NOMA. Our results lead to three key properties linking channel statistics and NOMA performance. In particular, correlated channels and similar channel statistics increase the variance. Further, correlated Rayleigh channels with a narrow angular spread yield the highest variance and have the highest cluster percentage, intra-cluster correlation and cluster-rate in comparison to other channel models. Therefore, we can conclude that correlated channels having similar channel statistics are most suitable for NOMA-ZFBF systems.

Chapter 7

Conclusion and Future Work

7.1 Conclusions

Current and future generation cellular communication must deal with:

1. an increasing demand for higher data rate;
2. dense user environments, which are likely to result in increased correlation between the channels of different users;
3. an increasing use of delay-constrained traffic, which raises concerns about fairness among users while scheduling the users for transmission.

Traditional linear signal processing schemes such as MF, MMSE and ZF are struggling to fulfill the above demands and provide spectral efficiency along with fairness among users. Analyzing the spectral efficiency performance and fairness of MU-MIMO systems when the users' channels are heterogeneous leads to a more thorough understanding of the performance of the current and future cellular networks. Therefore, this thesis analyzed and evaluated the performance of MU-MIMO systems with a ZF linear signal processing scheme, that provides higher rate along with providing fairness among the users. While doing so, this thesis presents two techniques to tackle this situation, which results from having a dense user environment, where both semi-orthogonal users and highly correlated users are present.

- Chapter 3 proposes a novel linear signal processing scheme that can increase the performance of the system when highly correlated users are present. The key contributions of this chapter are:

  - A novel and efficient DZF scheme has been proposed that can achieve higher performance in highly correlated user environments. DZF is a variant of ZF with very little additional complexity compared to CZF that provides robustness against the joint scheduling of correlated users. It focuses on gaining the most decorrelation with the least loss in power for highly correlated users.
  - The analysis presented in chapter 3 clearly demonstrates the idea of DZF and illustrates the potential for improved rates and fairness.

- On the basis of the DZF scheme, chapter 4 presents a HZF scheme. HZF is a technique for switching between CZF and DZF schemes based on the correlation threshold value of the scheduled users.

  - We have shown that HZF can be implemented by a simple yet efficient algorithm that provides fairness while scheduling the users and also achieves a performance gain.

- Chapter 5 focuses on a downlink NOMA and a power allocation strategy that ensures QoS for the weak user. This chapter also presents an analytical framework for characterizing the overall cluster-rate and sum-rate of the system. The key contributions of this chapter are as follows:

  - Considering a target SINR for the weak user, i.e., guaranteeing the weak user's QoS (W-QoS), a simple W-QoS based NOMA algorithm is proposed. The W-QoS NOMA algorithm checks the basic requirement for NOMA cluster formation and power allocation that always exceeds the TDMA rate.

- In order to analyze the effect of using the W-QoS NOMA with general channels, we integrate the proposed W-QoS NOMA with highly-correlated user clustering algorithms. Numerical results confirm that integrating the W-QoS NOMA and highly-correlated user clustering algorithms improves the sum-capacity of the system.
- Assuming perfectly correlated channels for the strong and weak users in each NOMA cluster, we derive the exact SNR/SINR distributions of singletons and strong and weak users in clusters in order to bound performance. Further, we investigate how close highly-correlated user clustering can get to this bound. From simulations, we observed that the analysis gives a tight upper bound. This is the only analysis of a NOMA-ZFBF system which includes power allocation, guarantees a NOMA rate is higher than TDMA rate and gives an exact tight upper bound for the sum-rate.

- The impact of dominant LoS directions and spatial correlation on the NOMA performance is investigated in Chapter 6.
  - Under the correlated Rician fading channel model, the variance of the interference power between two users is derived. This variance analysis proves that correlated channels with similar channel statistics increase the variance.
  - Overall, the results from this chapter demonstrate that an increase in variance means a more heterogeneous user environment which is likely to give good NOMA performance. Such a variance analysis is useful in order to understand the impact of different fading conditions on NOMA systems.

In this thesis we have provided two novel signal processing schemes and analyzed ZFBF based NOMA systems considering the highly correlated user environments that are likely to be present in the dense MU-MIMO systems of 5G cellular communication networks. The novel approach presented in Chapters 3 and 4 can serve as a basis for related research areas in future. In addition, this

thesis has contributed to enhance the understanding of NOMA-ZFBF based scheme through some analytical and numerical results in chapter 5 and channel models in chapter 6.

7.2 Future Work

It is known that 5G cellular systems will require several innovative solutions in comparison to existing 4G systems, some of which include adaptation of small cells, use of mmWave communications with short-range links, massive antenna arrays installed in macro BS and possibly deployment of cognitive radios to exploit the spectrum holes [122]. Thus, the investigations carried out and conclusions developed throughout the thesis lead us to propose some unexplored areas that are related to the contents of the thesis and are beneficial for 5G cellular communication.

7.2.1 Millimeter communication with DZF and NOMA

Since NOMA can support multiple users and mmWave provides a larger bandwidth, NOMA integrated with mmWave can efficiently utilize acquired spectrum to support massive connectivity. Furthermore, as mmWave frequency band propagation characteristic includes highly directional feature [11] [123] [124], this easily facilitates the integration of our high correlation based systems in mmWave communication system for closely spaced users. Moreover, to capture these physical properties of mmWave channels, cluster based spatial channel models, also known as Saleh-Valenzuela channel models, are often considered [14] [25] [55] [125] [126]. These channel models consider a finite number of scatterers and paths which depends upon the random geometry of the propagation environment. This random propagation environment along with transmit and receive antenna array responses makes it more difficult to make analytical progress. However it can be numerically simulated to gain a proper understanding of their practicality. Naturally, some of the properties of dominant parameters, such as angular spread and LoS angle, can be easily understood and some analytical progress can be obtained.

Therefore, as studied in chapters 3 - 6, both DZF and NOMA-ZFBF schemes can be further extended and analyzed with cluster based spatial channel models. If some analytical results can be obtained then it will lead to further potential for design and optimization of transceiver signal processing schemes. This can be used as a basis for design and system performance evaluation of mmWave based cellular systems. Moreover, in chapter 6 we have studied the channel correlation using transmit correlation in a one ring model. Therefore, including receive correlation can also be a topic of future work [127].

7.2.2 Imperfect channel state information

As discussed in chapter 2.3.1, the ZF precoding technique is dependent upon the channel state information at transmitter (CSIT). Perfect CSIT being available at the receiver means more precise interference cancellation. However, due to the time varying nature of wireless channels as well as the limited resource available for channel estimation, perfect CSIT is not possible [128]. Therefore, a study of NOMA systems considering imperfect CSI would provide more understanding of the reliability of NOMA systems [27]. In NOMA-ZFBF systems two or more users are sharing the same beam and the precoder is only designed on the basis of the strong user's channel information which results in imperfect precoding information for the weak user, unless the weak and strong users are perfectly correlated. Therefore, it is important to design systems with imperfect CSIT, that require less feedback and also to try and improve the weak user performance.

7.2.3 Mixed beam NOMA

In chapters 5 and 6, ZFBF based NOMA systems are discussed and the precoders for clusters are designed based on the strong user channels. Therefore, the weak user suffers from interference from other clusters and from their strong user partner. A novel approach to optimize the precoder design and the overall system could use an extra "*mixing*" parameter. *Mixing* can be done while designing a beamformer in a

beam or by using a mixture of channels of both users in a cluster. For example, denote $\mathbf{w}_i^s$ and $\mathbf{w}_i^w$ as the normalized beamformers for the near/strong and weak/far users in cluster i. The unnormalized mixed beamformer can be computed as $\tilde{\mathbf{w}}_i = r_i\mathbf{w}_i^s + (\sqrt{1-r_i^2})\mathbf{w}_i^w$ and the final beamformer can be obtained by using $\mathbf{w}_i = \frac{\tilde{\mathbf{w}}_i}{||\tilde{\mathbf{w}}_i||}$. Similarly, we can use a mixture of channels as $r_i\mathbf{H}_i^s + (\sqrt{1-r_i^2})\mathbf{H}_i^w$ in clusters and design precoders as in (3.3) and (3.4). Performance evaluation in terms of complexity and rate would be interesting.

7.2.4 Quasi-degradation concept in NOMA

We know that ZFBF approaches are effective to improve system throughput in both OMA and NOMA, where spatial degrees of freedom are used for interference avoidance. ZFBF mitigates multi-user interference in OMA or inter beam interference in NOMA, by transmitting data in the null space of other users' channel matrices [76] [129]. However, this approach is efficient when the number of transmit antennas are not smaller than the number of receive antennas. Recently, the concept of quasi-degraded channels are being developed for heterogeneous broadcast system. [130] used quasi-degradation to characterize the gap between the optimal performance of DPC compared with that of NOMA. The obtained solution considering a heterogeneous case with distinctive users' channel conditions suggest that the performance gap is zero. Further, in-depth study has been undertaken in [129] to illustrate important properties of the quasi-degradation concept. Thus, as suggested in [129], the future direction can be to study the NOMA system design with quasi-degraded channels.

Bibliography

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101 – 107, 2011.

[2] "SK Telecom view in 5G vision, architecture, technology, and spectrum," *SK Telecom 5G White Paper, Network Technology R & D Centre*, 2014.

[3] "3GPP TS 36.300, evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description," available: https://portal.3gpp.org.

[4] "3GPP TR 36.814, evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects; overall description," available: https://portal.3gpp.org.

[5] N. Himayat, S. Talwar, A. Rao, and R. Soni, "Interference management for 4G cellular standards [WIMAX/LTE UPDATE]," *IEEE Commun. Mag.*, vol. 48, no. 8, pp. 86–92, 2010.

[6] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *Proc. Intelligent Signal Process. and Commun. Systems (ISPACS)*, 2013, pp. 770–774.

[7] DOCOMO, "5G radio access: Requirements, concepts and technologies," *DOCOMO 5G White Paper, NTT DOCOMO, INC.*

[8] METIS, "Mobile and wireless communications enablers for the twenty-twenty information society (METIS)," *Document Number: ICT-317669-METIS/D2.1*, 2014.

[9] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[10] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, 2014.

[11] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5G: Opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, 2015.

[12] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[13] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, 2014.

[14] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. on Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[15] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, 2013.

[16] A. I. Sulyman, A. T. Nassar, M. K. Samimi, G. R. MacCartney, T. S. Rappaport, and A. Alsanie, "Radio propagation path loss models for 5G cellular networks in the 28 GHz and 38 GHz millimeter-wave bands," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 78–86, 2014.

[17] G. R. MacCartney, J. Zhang, S. Nie, and T. S. Rappaport, "Path loss models for 5G millimeter wave propagation channels in urban microcells," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2013, pp. 3948–3953.

[18] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng, "Indoor office wideband millimeter-wave propagation measurements and channel models at 28 and 73 GHz for ultra-dense 5G wireless networks," *IEEE Access*, vol. 3, pp. 2388–2424, 2015.

[19] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE J. on Sel. Areas Commun.*, vol. 5, no. 5, pp. 871–878, 1987.

[20] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[21] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Trans. on Emerging Telecommun. Technologies*, vol. 10, no. 6, pp. 585–595, 1999.

[22] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, 1998.

[23] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

[24] "3GPP TR 36.873 V12.3.0: Study on 3D channel model for LTE (release 12)," December 2016, available: https://3gpp.org.

[25] "3GPP TR 38.900 V14.3.1: Study on channel model for frequency spectrum above 6 GHz (release 14)," Apr. 2017, available: https://portal.3gpp.org.

[26] Z. Ding, F. Adachi, and H. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, 2016.

[27] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, 2017.

[28] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. New York, USA: Cambridge University Press, 2007.

[29] G. Tsoulos, *MIMO System Technology for Wireless Communications*. Boca Raton, FL, USA: CRC press, 2006.

[30] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, 2005.

[31] C. Wang, E. K. Au, R. D. Murch, W. H. Mow, R. S. Cheng, and V. Lau, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 3, pp. 805 – 810, 2007.

[32] R. W. Heath, M. Airy, and A. J. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," in *Proc. Asilomar Conference on Signals, Systems and Computers*, vol. 2. IEEE, 2001, pp. 1194–1199.

[33] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. on Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, 2006.

[34] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. New York, NY, USA: Cambridge University Press, 2003.

[35] D. A. Gore, R. W. Heath, and A. J. Paulraj, "Transmit selection in spatial multiplexing systems," *IEEE Commun. Lett.*, vol. 6, no. 11, pp. 491–493, 2002.

[36] F. Kaltenberger, D. Gesbert, R. Knopp, and M. Kountouris, "Correlation and capacity of measured multi-user MIMO channels," in *Proc. Personal, Indoor, and Mobile Radio Communication (PIMRC)*. IEEE, 2008, pp. 1–5.

[37] M. Kiessling and J. Speidel, "Analytical performance of MIMO zero-forcing receivers in correlated Rayleigh fading environments," in *Proc. IEEE Signal Process. Advances in Wireless Commun. (SPAWC)*, 2003, pp. 383–387.

[38] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE Veh. Tech. Conf. (VTC Fall)*, 2011, pp. 1–5.

[39] L. Sun and M. R. McKay, "Eigen-based transceivers for the MIMO broadcast channel with semi-orthogonal user selection," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5246–5261, 2010.

[40] B. Kim, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. Military Commun. Conf. (MILCOM)*, 2013, pp. 1278–1283.

[41] J. Flordelis, X. Gao, G. Dahman, F. Rusek, O. Edfors, and F. Tufvesson, "Spatial separation of closely-spaced users in measured massive multi-user MIMO channels," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, 2015, pp. 1441–1446.

[42] Z. Xiao, L. Dai, Z. Ding, J. Choi, and P. Xia, "Millimeter-wave communication with non-orthogonal multiple access for 5G," *arXiv preprint arXiv:1709.07980*, 2017.

[43] Y. Saito, A. Benjebbour, Y. Kishiyama, X. Wang, X. Hou, H. Jiang, L. Lu, W. Liang, B. Li, L. Gu, Y. Cui, and T. Kashima, "Large scale field experimental trial of downlink TDD massive MIMO at the 4.5 GHz band," in *Proc. IEEE Veh. Tech. Conf. (VTC Spring)*, June 2017, pp. 1–5.

[44] T. Kashima, J. Qiu, H. Shen, C. Tang, T. Tian, X. Wang, X. Hou, H. Jiang, A. Benjebbour, Y. Saito *et al.*, "Large scale massive MIMO field trial for 5G mobile communications system," in *Proc. IEEE Int. Symp. on Antennas and Propagation (ISAP)*, 2016, pp. 602–603.

[45] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.

[46] C. Jinho, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 791–800, 2015.

[47] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 2, pp. 721–742, 2017.

[48] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[49] Z. Gao, L. Dai, C. Yuen, and Z. Wang, "Asymptotic orthogonality analysis of time-domain sparse massive MIMO channels," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1826–1829, 2015.

[50] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[51] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[52] W. A. W. M. Mahyiddin, "Training in massive MIMO systems," Ph.D. dissertation, University of Canterbury, 2015.

[53] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, 2016.

[54] D. Basnayaka, "Macrodiversity MIMO transceivers," Ph.D. dissertation, University of Canterbury, 2012.

[55] H. Tataria, "Analysis of multiuser cellular systems over heterogeneous channels," Ph.D. dissertation, Victoria University of Wellington, 2017.

[56] R. van Nobelen, "Coding for the Rayleigh fading channel," Ph.D. dissertation, University of Canterbury, 1996.

[57] A. F. Molisch, *Wireless Communications*. John Wiley & Sons, 2011.

[58] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Mordock, *Millimeter Wave Wireless Communications*. Pearson Education, 2014.

[59] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. on Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[60] W. Wang, A. Harada, and H. Kayama, "Enhanced limited feedback schemes for DL MU-MIMO ZF precoding," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 4, pp. 1554–1561, 2013.

[61] C. T. Neil, "On the performance and analysis of massive MIMO or 5G wireless systems," Ph.D. dissertation, Victoria University of Wellington, 2017.

[62] J. G. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2001.

[63] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. on Commun.*, vol. 48, no. 3, pp. 502–513, 2000.

[64] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *Proc. IEEE Personal Commun.*, vol. 5, no. 1, pp. 10–22, 1998.

[65] P. Petrus, J. H. Reed, and T. S. Rappaport, "Geometrically based statistical channel model for macrocellular mobile environments," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, vol. 2. IEEE, 1996, pp. 1197–1201.

[66] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Systems Journal*, pp. 1–12, 2017.

[67] X. Gao, "Massive MIMO in real propagation environments," Ph.D. dissertation, Lund University, 2016.

[68] Z. Ge and W. Haiyan, "Linear precoding design for massive MIMO based on the minimum mean square error algorithm," *EURASIP J. on Embedded Systems*, vol. 2017, no. 1, p. 20, 2017.

[69] J.-C. Chen, C.-J. Wang, K.-K. Wong, and C.-K. Wen, "Low-complexity precoding design for massive multiuser MIMO systems using approximate message passing," *IEEE Trans. on Vehicular Tech.*, vol. 65, no. 7, pp. 5707–5714, 2016.

[70] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. on Info. Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[71] R. Doostnejad, "Precoding and beamforming for multi-input multi-output downlink channels," Ph.D. dissertation, University of Toronto, 2005.

[72] C. Windpassinger, R. F. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Trans. on Commun.*, vol. 52, no. 12, pp. 2057–2060, 2004.

[73] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. of Sel. Topics in Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[74] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. on Signal Process.*, vol. 56, no. 9, pp. 4409–4418, 2008.

[75] A. Tarighat, M. Sadek, and A. H. Sayed, "A multi user beamforming scheme for downlink MIMO channels based on maximizing signal-to-leakage ratios," in *Proc. IEEE International Conference Acoustics, Speech, and Signal Process. (ICASSP'05)*, vol. 3, 2005, pp. III–1129.

[76] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb 2004.

[77] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[78] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," in *Proc. IEEE Int. Symp. on Wireless Commun. Systems (ISWCS)*, 2014, pp. 781–785.

[79] O. Nagisa, Y. Kishiyama, and K. Higuchi, "Performance of non-orthogonal multiple access with SIC in cellular downlink using proportional fair-based resource allocation," *IEICE Trans. Commun.*, vol. 98, no. 2, pp. 344–351, 2015.

[80] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE Veh. Tech. Conf. (VTC Spring)*, 2013, pp. 1–5.

[81] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, 2014.

[82] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, 2015.

[83] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Annu. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, 2013, pp. 611–615.

[84] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, 2014.

[85] C. Xiaohang, A. Benjebbour, L. Anxin, and A. Harada, "Multi-user proportional fair scheduling for uplink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, 2014, pp. 1–5.

[86] K. Beomju, C. Wonsuk, L. Sungmook, S. Sangwook, K. Jonghyung, C. Sooyong, and H. Daesik, "Uplink NOMA with multi-antenna," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, 2015, pp. 1–5.

[87] A. Li, A. Benjebbour, and A. Harada, "Performance evaluation of non-orthogonal multiple access combined with opportunistic beamforming," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, 2014, pp. 1–5.

[88] Y. Hayashi, Y. Kishiyama, and K. Higuchi, "Investigations on power allocation among beams in non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink," in *Proc. IEEE Veh. Tech. Conf. (VTC Fall)*, 2013, pp. 1–5.

[89] Y. Chunlin, A. Harada, A. Benjebbour, L. Yang, L. Anxin, and J. Huiling, "Receiver design for downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, 2015, pp. 1–6.

[90] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "System-level performance of downlink NOMA for future LTE enhancements," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2013, pp. 66–70.

[91] N. Nonaka, A. Benjebbour, and K. Higuchi, "System-level throughput of NOMA using intra-beam superposition coding and SIC in MIMO downlink when channel estimation error exists," in *Proc. IEEE Int. Conf. Commun. Systems (ICCS)*, 2014, pp. 202–206.

[92] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2628–2639, 2012.

[93] H. Haci, "Non-orthogonal multiple access (NOMA) with asynchronous interference cancellation," Ph.D. dissertation, University of Kent, 2015.

[94] J. G. Andrews, "Interference cancellation for cellular systems: a contemporary overview," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 19–29, 2005.

[95] N. I. Miridakis and D. D. Vergados, "A survey on the successive interference cancellation performance for single-antenna and multiple-antenna OFDM systems," *IEEE Commun. Surveys and Tutorials*, vol. 15, no. 1, pp. 312–335, 2013.

[96] S. Liu, C. Zhang, and G. Lyu, "User selection and power schedule for downlink non-orthogonal multiple access (NOMA) system," in *Proc. ICCW*, 2015, pp. 2561–2565.

[97] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. on Commun.*, vol. 98, no. 3, pp. 403–414, 2015.

[98] Y. J. Zhang and K. B. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Trans. on Commun.*, vol. 53, no. 1, pp. 107–116, 2005.

[99] A. B. Sediq, R. H. Gohary, and H. Yanikomeroglu, "Optimal tradeoff between efficiency and Jain's fairness index in resource allocation," in *Proc. Personal, Indoor, and Mobile Radio Communication (PIMRC)*, 2012, pp. 577–583.

[100] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.

[101] D. J. Love, R. W. Heath, V. K. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, 2008.

[102] C. T. Neil, M. Shafi, P. J. Smith, P. A. Dmochowski, and J. Zhang, "Impact of microwave and mmWave channel models on 5G systems performance," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6505–6520, 2017.

[103] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, 2016.

[104] H. Tataria, P. J. Smith, L. J. Greenstein, and P. A. Dmochowski, "Zero-forcing precoding performance in multiuser MIMO systems with heterogeneous Ricean fading," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 74–77, 2017.

[105] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Magazine*, vol. 53, no. 9, pp. 74–81, 2015.

[106] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *arXiv preprint arXiv:1706.05347*, 2017.

[107] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. 98, no. 3, pp. 403–414, 2015.

[108] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, 2016.

[109] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 84–87, 2017.

[110] Y. Zhang, H.-M. Wang, Q. Yang, and Z. Ding, "Secrecy sum rate maximization in non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 930–933, 2016.

[111] J. Kim, J. Koh, J. Kang, K. Lee, and J. Kang, "Design of user clustering and precoding for downlink non-orthogonal multiple access (NOMA)," in *Proc. Military Commun. Conf. (MILCOM)*, 2015, pp. 1170–1175.

[112] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, 2015.

[113] Q. Sun, S. Han, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Commun. Lett.*, vol. 4, no. 4, pp. 405–408, 2015.

[114] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions." *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, 2016.

[115] P. Xu, Z. Ding, X. Dai, and H. V. Poor, "NOMA: An information theoretic perspective," *arXiv preprint arXiv:1504.07751*, 2015.

[116] Y. Zhao, M. Zhao, S. Zhou, and J. Wang, "Closed-form capacity expressions for SIMO channels with correlated fading," in *Proc. IEEE Veh. Tech. Conf. (VTC Fall)*, vol. 2, pp. 982–985.

[117] Y. Liu, G. Pan, H. Zhang, and M. Song, "On the capacity comparison between MIMO-NOMA and MIMO-OMA," *IEEE Access*, vol. 4, pp. 2123–2129, 2016.

[118] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access NOMA," in *Proc. Personal, Indoor, and Mobile Radio Communication (PIMRC)*, 2013, pp. 611–615.

[119] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "System-level performance of downlink NOMA for future LTE enhancements," in *Globecom Workshops (GC Wkshps)*. IEEE, 2013, pp. 66–70.

[120] R. Jiao, L. Dai, J. Zhang, R. MacKenzie, and M. Hao, "On the performance of NOMA-based cooperative relaying systems over Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11 409–11 413, 2017.

[121] H. Tataria, P. J. Smith, L. J. Greenstein, P. A. Dmochowski, and M. Matthaiou, "Impact of line-of-sight and unequal spatial correlation on uplink MU-MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 634–637, 2017.

[122] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 80–93, 2014.

[123] J. Bae, Y. S. Choi, J. S. Kim, and M. Y. Chung, "Architecture and performance evaluation of mmwave based 5G mobile communication system," in *Proc. Int. Conf. on Info. and Commun. Tech. Convergence (ICTC)*. IEEE, 2014, pp. 847–851.

[124] T. E. Bogale and L. B. Le, "Massive MIMO and mmwave for 5G wireless hetnet: Potential benefits and challenges," *IEEE Veh. Tech. Mag.*, vol. 11, no. 1, pp. 64–75, 2016.

[125] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, 2016.

[126] ——, "3-D statistical channel model for millimeter-wave outdoor mobile broadband communications," in *Proc. IEEE Int. Conf. on Commun. (ICC)*. IEEE, 2015, pp. 2430–2436.

[127] B. Nosrat-Makouei, J. G. Andrews, and R. W. Heath, "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Trans. on Signal Process.*, vol. 59, no. 6, pp. 2783–2794, 2011.

[128] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, 2013.

[129] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to miso-noma downlink," *IEEE Trans. on Signal Process.*, vol. 64, no. 23, pp. 6174–6189, 2016.

[130] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a qos optimization problem in two-user miso-noma downlink," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1263–1266, 2016.

Appendix A

A.1 HZF- Exhaustive search method

Table A.1: Combination of 2 users in HZF Exhaustive search method

User Rates	User 1	User 2
R1 (CZF)	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(1)}$
R2	$\mathbf{u}_1^{(2)}$	$\mathbf{u}_2^{(1)}$
R3	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(2)}$

Table A.2: Combination of 3 users in HZF Exhaustive search method

User Rates	User 1	User 2	User 3
R1 (CZF)	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(1)}$
R2	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(2)}$
R3	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(2)}$	$\mathbf{u}_3^{(1)}$
R3	$\mathbf{u}_1^{(2)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(1)}$

In tables A.1, A.2 and A.3, $\mathbf{u}_i^{(1)}$ means first eigen vector is used for receiving processing to create effective channels. Similarly, $\mathbf{u}_i^{(2)}$ means second eigen vector is used.

Table A.3: Combination of 4 users in HZF Exhaustive search method

User Rates	User 1	User 2	User 3	User 4
R1 (CZF)	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(1)}$	$\mathbf{u}_4^{(1)}$
R2	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(1)}$	$\mathbf{u}_4^{(2)}$
R3	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(2)}$	$\mathbf{u}_4^{(1)}$
R4	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(2)}$	$\mathbf{u}_3^{(1)}$	$\mathbf{u}_4^{(1)}$
R5	$\mathbf{u}_1^{(2)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(1)}$	$\mathbf{u}_4^{(1)}$
R6	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(2)}$	$\mathbf{u}_4^{(2)}$
R7	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(2)}$	$\mathbf{u}_3^{(1)}$	$\mathbf{u}_4^{(2)}$
R8	$\mathbf{u}_1^{(1)}$	$\mathbf{u}_2^{(2)}$	$\mathbf{u}_3^{(2)}$	$\mathbf{u}_4^{(1)}$
R9	$\mathbf{u}_1^{(2)}$	$\mathbf{u}_2^{(2)}$	$\mathbf{u}_3^{(1)}$	$\mathbf{u}_4^{(1)}$
R10	$\mathbf{u}_1^{(2)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(2)}$	$\mathbf{u}_4^{(1)}$
R11	$\mathbf{u}_1^{(2)}$	$\mathbf{u}_2^{(1)}$	$\mathbf{u}_3^{(1)}$	$\mathbf{u}_4^{(2)}$

Appendix B

B.1 Proof of Cluster rate

The rate for a cluster depends on whether the scheduled cluster supports a pair of users or if the weak user was dropped leading to a singleton. If $Y_i > Y_0$ then a NOMA cluster is possible and the rate equals the sum of the strong and weak user rates. On the other hand, if $Y_i \leq Y_0$ then the cluster is dropped. Hence,

$$E[R_{c,i}] = P(Y_i > Y_0)E[R_{c,i}|Y_i > Y_0] + P(Y_i \leq Y_0)E[R_{c,i}|Y_i \leq Y_0]. \tag{B.1}$$

The first conditional mean rate in (B.1) can be written as

$$E[R_{c,i}\big|Y_i > Y_0] = E\bigg[\log\Big(1 + \rho\beta_{is}Y_i - T(\frac{\beta_{is}}{\beta_{iw}} - 1)\Big)\bigg|Y_i > Y_0\bigg]. \tag{B.2}$$

Similarly, the second conditional mean rate can be written as

$$E[R_{c,i}\big|Y_i \leq Y_0] = E\bigg[\log\Big(1 + \rho\beta_{is}Y_i\Big)\bigg|Y_i \leq Y_0\bigg]. \tag{B.3}$$

Now, equation (B.2) can be written as

$$\begin{aligned} E\bigg[\log\Big(a + bY\Big)\Big|Y > Y_0\bigg] &= \int_{Y_0}^{\infty} \frac{\log(a+bx)f_p(x)}{1 - F_p(Y_0)}dx \\ &= \frac{I_1(a,b)}{1 - F_p(Y_0)}, \end{aligned} \tag{B.4}$$

where $f_p(x) = F_p'(x)$ is the χ^2_p PDF. Substituting $f_p(x)$ in (B.4) and using the change of variable $y = x - Y_0$ gives

$$I_1(a,b) = \int_0^\infty \log\big(a + b(Y_0 + y)\big) \frac{(y+Y_0)^{p-1}}{(p-1)!} e^{-(y+Y_0)} dy. \tag{B.5}$$

Expanding $(y + Y_0)^{p-1}$ and using standard log results gives

$$I_1(a,b) = \frac{1}{(p-1)!} \sum_{q=0}^{p-1} \binom{p-1}{q} Y_0^{p-1-q} e^{-Y_0} \times \int_0^\infty \Big(\log(a + bY_0) + \log(1 + cy)\Big) y^q e^{-y} dy, \tag{B.6}$$

where $c = b(a + bY_0)^{-1}$. Using known integrals and

$$E[\log(1 + \rho\beta_i Y_i] = C(p, \rho\beta_i), \tag{B.7}$$

(B.6) becomes

$$I_1(a,b) = \frac{1}{(p-1)!} \sum_{q=0}^{p-1} \binom{p-1}{q} Y_0^{p-1-q} e^{-Y_0} \times \left\{ \log(a + bY_0) q! + q!\, C(q+1, \frac{b}{a + bY_0}) \right\}. \tag{B.8}$$

Similarly,

$$E[R_{c,i} \big| Y_i \le Y_0] = E\left[\log\Big(1 + \rho\beta_{is} Y_i\Big) \Big| Y_i \le Y_0 \right]. \tag{B.9}$$

can be rewritten as

$$\begin{aligned} E\left[\log\Big(a + bY\Big) \Big| Y \le Y_0 \right] &= \int_0^{Y_0} \frac{\log(a + bx) f_p(x)}{F_p(Y_0)} dx \\ &= \frac{I_2}{F_p(Y_0)}. \end{aligned} \tag{B.10}$$

From (B.7) we can rewrite (B.10) as

$$I_2 = C(p,b) - \int_{Y_0}^\infty \log(a + bx) f_p(x) dx. \tag{B.11}$$

Noting that the integral in (B.11) is a special case of I_1 with $a = 1$, we immediately obtain

$$I_2 = C(p, b) - I_1(1, b). \tag{B.12}$$

Substituting $a = 1 - T\Big(\frac{\beta_{is}}{\beta_{iw}} - 1\Big)$, $b = \rho\beta_{is}$ in (B.8) and (B.12) and substituting I_1 and I_2 in (5.56) gives the desired result.

$$E[R_{c,i}] = I_1(a_i, b_i) + C(p, b_i) - I_1(1, b_i), \tag{B.13}$$

where $a_i = 1 - T\Big(\frac{\beta_{is}}{\beta_{iw}} - 1\Big)$, $b_i = \rho\beta_{is}$, $C(\cdot, \cdot)$ is [116, Eq. 7] and $I_1(\cdot, \cdot)$ is the closed form function given in (B.8).

B.2 SUS-CUC algorithm

The SUS-CUC algorithm is given below.

1. Initialization stage:

   $\Xi_1 = \{1, ..., L\}$, $i = 1$, $\gamma = \emptyset$, $\tau_c = \{\emptyset\}$, $\tau_d = \emptyset$,

   where L is the total number of participating users, γ is the set of orthogonal users, τ_c is the set of clusters and τ_d is the set of dropped users.

2. SUS algorithm [33]: Select a semi-orthogonal user $\pi(i)$ and the set of users, Ξ_{i+1}, that are semi-orthogonal with $\pi(i)$.

3. CUC algorithm:

   (a) Set $\tau_d \leftarrow \tau_d \cup \{\Xi_i - \Xi_{i+1} - \pi(i)\}$.

   (b) Calculate the set of correlated users to the $\pi(i)$ user from the set of dropped users, τ_d, as

   $$\varphi = \left\{ l \in \tau_d, \Big| \frac{|\mathbf{h}_l \mathbf{g}^*_{(i)}|}{||\mathbf{h}_l|| ||\mathbf{g}_{(i)}||} > \Psi \right\}, \tag{B.14}$$

   where Ψ is the correlation threshold.

(c) This step depends on the value of φ from step (b).

- If $|\varphi| > 1$, then for each user in φ, the transmitter calculates the channel gain between these users and orthogonal user $\pi(i)$ and selects the user pair having maximum absolute channel gain difference. $\nu = \arg\max_{j \in \varphi} d_{(\pi(i),j)} = \Big| ||\mathbf{h}_{\pi(i)}|| - ||\mathbf{h}_j|| \Big|$,
  $\tau_c \leftarrow \{\tau_c\} \cup \{\tau_{c(i)}\}$, where $\tau_{c(i)} = \{\pi(i), \nu\}$, $\tau_d \leftarrow \tau_d - \{\nu\}$,
- If $|\varphi| = 1$, $\tau_c \leftarrow \{\tau_c\} \cup \{\tau_{c(i)}\}$, where $\tau_{c(i)} = \{\pi(i), \nu\}$ and update τ_d as $\tau_d \leftarrow \tau_d - \{\nu\}$.
- If $|\varphi| = 0$, there are no highly correlated users and the cluster $\tau_{c(i)}$ contains only singleton user $\pi(i)$ as $\tau_{c(i)} = \{\pi(i), \emptyset\}$. Update $\tau_d \leftarrow \tau_d$.

(d) If $|\gamma| \leq M$, then go to step 2. Otherwise, go to step 4.

4. Order paired users according to their channel powers.

In SUS-CUC, once the best user $\pi(i)$ is selected and all the non-orthogonal users are dropped off, SUS-CUC immediately starts to find the best pair for that user $\pi(i)$. In step 3(b) of SUS-CUC, which is the correlated user pairing part, we consider g_i, the component of h_i orthogonal to the subspace spanned by $\{g_{(1)}, \cdots, g_{(i-1)}\}$ as described in step 2 of the SUS algorithm in [33]. From the SUS-CUC algorithm we obtain two types of clusters in τ_c as: a) clusters with two (strong and weak) users ($\tau_{c(i)} = \{\pi(i), \nu\}$) namely NOMA-clusters and b) clusters with one user ($\tau_{c(i)} = \{\pi(i), \emptyset\}$) namely singletons.

Appendix C

C.1 Detailed derivation of variance

The spatially correlated Rician fading channel can be written as

$$\mathbf{g}_i = \sqrt{\frac{K_i}{K_i+1}}\ \bar{\mathbf{h}}_i + \sqrt{\frac{1}{K_i+1}}\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2}, \tag{C.1}$$

The Rician channel model in (C.1) can be written as

$$\mathbf{g}_i = \mathbf{a}_i + b_i\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2}. \tag{C.2}$$

Expanding $\mathbf{g}_i \mathbf{g}_j^H$ gives

$$\begin{aligned} \mathbf{g}_i \mathbf{g}_j^H &= \left(\mathbf{a}_i + b_i\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2}\right)\left(\mathbf{a}_j + b_j\ \hat{\mathbf{h}}_j \mathbf{R}_j^{1/2}\right)^H \\ &= \mathbf{a}_i \mathbf{a}_j^H + \mathbf{a}_i \mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H b_j + b_i\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2} \mathbf{a}_j^H + b_i b_j\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2} \mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H. \end{aligned} \tag{C.3}$$

Then,

$$\begin{aligned} |\mathbf{g}_i \mathbf{g}_j^H|^2 = &\left(\mathbf{a}_i \mathbf{a}_j^H + \mathbf{a}_i \mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H b_j + b_i\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2} \mathbf{a}_j^H + b_i b_j\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2} \mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H\right) \\ &\times \left(\mathbf{a}_i \mathbf{a}_j^H + \mathbf{a}_i \mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H b_j + b_i\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2} \mathbf{a}_j^H + b_i b_j\ \hat{\mathbf{h}}_i \mathbf{R}_i^{1/2} \mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H\right)^H. \end{aligned} \tag{C.4}$$

Taking expectation of (C.4),

$$E\big[|\mathbf{g}_i\mathbf{g}_j^H|^2\big] = E\big[\mathbf{a}_i\mathbf{a}_j^H\mathbf{a}_j\mathbf{a}_i^H + \mathbf{a}_i\mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H b_j^2\ \hat{\mathbf{h}}_j\mathbf{R}_j^{1/2}\mathbf{a}_i^H + b_i\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}\mathbf{a}_j^H b_i\mathbf{a}_j\mathbf{R}_i^{1/2}\ \hat{\mathbf{h}}_i^H + b_i^2 b_j^2\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}\mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H\ \hat{\mathbf{h}}_j\mathbf{R}_j^{1/2}\mathbf{R}_i^{1/2}\ \hat{\mathbf{h}}_i^H\big]. \tag{C.5}$$

Using the known results

$$E[\ \hat{\mathbf{h}}_i] = \mathbf{0}, \tag{C.6a}$$

$$E[\ \hat{\mathbf{h}}_i\ \hat{\mathbf{h}}_j^H] = 0, \tag{C.6b}$$

$$E[\ \hat{\mathbf{h}}_i^H\ \hat{\mathbf{h}}_i] = \mathbf{I}, \tag{C.6c}$$

$$E[\ \hat{\mathbf{h}}_i\ \hat{\mathbf{h}}_i^H] = N, \tag{C.6d}$$

and

$$E[\ \hat{\mathbf{h}}_i\mathbf{A}\ \hat{\mathbf{h}}_i^H] = \mathrm{tr}(\mathbf{A}), \tag{C.6e}$$

(C.5) can be written as

$$\begin{aligned} E\big[|\mathbf{g}_i\mathbf{g}_j^H|^2\big] &= |\mathbf{a}_i\mathbf{a}_j^H|^2 + b_j^2\,\mathrm{tr}(\mathbf{R}_j^{1/2}\mathbf{a}_i^H\mathbf{a}_i\mathbf{R}_j^{1/2}) + b_i^2\,\mathrm{tr}(\mathbf{R}_i^{1/2}\mathbf{a}_j^H\mathbf{a}_j\mathbf{R}_i^{1/2}) + b_i^2 b_j^2\,\mathrm{tr}(\mathbf{R}_i^{1/2}\mathbf{R}_j\mathbf{R}_i^{1/2}) \\ &= |\mathbf{a}_i\mathbf{a}_j^H|^2 + b_j^2 \times (\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H) + b_i^2 \times (\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H) + b_i^2 b_j^2 \times (\mathbf{R}_i\mathbf{R}_j). \end{aligned} \tag{C.7}$$

Now $\mathrm{Var}\big[|\mathbf{g}_i\mathbf{g}_j^H|^2\big]$ needs

$$E\big[|\mathbf{g}_i\mathbf{g}_j^H|^4\big] = E\big[(\mathbf{g}_i\mathbf{g}_j^H\mathbf{g}_j\mathbf{g}_i^H)^2\big]. \tag{C.8}$$

Replacing $\mathbf{g}_j^H\mathbf{g}_j = \mathbf{Q}$, and with $\mathbf{g}_i$ from (C.2), (C.8) can be rewritten as

$$\begin{aligned} E\big[(\mathbf{g}_i\mathbf{Q}\mathbf{g}_i^H)^2\big] &= E\left[\left(\Big(\mathbf{a}_i + b_i\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}\Big)\mathbf{Q}\Big(\mathbf{a}_i^H + b_i\mathbf{R}_i^{1/2}\ \hat{\mathbf{h}}_i^H\Big)\right)^2\right] \\ &= E\left[\Big(\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H + \mathbf{a}_i\mathbf{Q}b_i\mathbf{R}_i^{1/2}\ \hat{\mathbf{h}}_i^H + b_i\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}\mathbf{Q}\mathbf{a}_i^H + b_i\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}\mathbf{Q}b_i\mathbf{R}_i^{1/2}\ \hat{\mathbf{h}}_i^H\Big)^2\right] \end{aligned} \tag{C.9}$$

$$E\big[(\mathbf{g}_i\mathbf{Q}\mathbf{g}_i^H)^2\big] = E\Big[\Big(\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H\Big)^2\Big] + 2E\Big[\Big(\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H b_i\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}\mathbf{Q}b_i\mathbf{R}_i^{1/2}\mathbf{a}_i^H\Big)\Big] + 2E\Big[\Big(\mathbf{a}_i\mathbf{Q}b_i\mathbf{R}_i^{1/2}\ \hat{\mathbf{h}}_i^H\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}\mathbf{Q}\mathbf{a}_i^H b_i\Big)\Big] + E\Big[\Big(b_i\ \hat{\mathbf{h}}_i\mathbf{R}_i^{1/2}\mathbf{Q}b_i\mathbf{R}_i^{1/2}\ \hat{\mathbf{h}}_i^H\Big)^2\Big]. \tag{C.10}$$

Using the results

$$E[\ \hat{\mathbf{h}}_i^H\ \hat{\mathbf{h}}_i\mathbf{A}\ \hat{\mathbf{h}}_i^H\ \hat{\mathbf{h}}_i] = \mathbf{A} + \mathrm{tr}(\mathbf{A})\mathbf{I}, \tag{C.11a}$$

and

$$E[(\ \hat{\mathbf{h}}_i\mathbf{A}\ \hat{\mathbf{h}}_i^H)^2] = \mathrm{tr}(\mathbf{A}^2) + (\mathrm{tr}(\mathbf{A}))^2, \tag{C.11b}$$

in addition to (C.6) gives

$$\begin{aligned}
E\big[(\mathbf{g}_i\mathbf{Q}\mathbf{g}_i^H)^2\big] &= E\Big[\Big(\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H\Big)^2\Big] + 2E\Big[\Big(\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H b_i^2\ \mathrm{tr}(\mathbf{R}_i^{1/2}\mathbf{Q}\mathbf{R}_i^{1/2})\Big)\Big] \\
&\quad + 2E\Big[\Big(b_i^2\mathbf{a}_i\mathbf{Q}\mathbf{R}_i^{1/2}\mathbf{R}_i^{1/2}\mathbf{Q}\mathbf{a}_i^H\Big)\Big] + b_i^4 E\Big[\mathrm{tr}[(\mathbf{R}_i^{1/2}\mathbf{Q}\mathbf{R}_i^{1/2})^2] + \mathrm{tr}\big((\mathbf{R}_i^{1/2}\mathbf{Q}\mathbf{R}_i^{1/2})\big)^2\Big] \\
&= E\Big[\Big(\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H\Big)^2\Big] + 2b_i^2 E\Big[\Big(\mathbf{a}_i\mathbf{Q}\mathbf{a}_i^H\ \mathrm{tr}(\mathbf{Q}\mathbf{R}_i)\Big)\Big] \\
&\quad + 2b_i^2 E\Big[\Big(\mathbf{a}_i\mathbf{Q}\mathbf{R}_i\mathbf{Q}\mathbf{a}_i^H\Big)\Big] + b_i^4 E\Big[\mathrm{tr}(\mathbf{Q}\mathbf{R}_i\mathbf{Q}\mathbf{R}_i) + \mathrm{tr}\big(\mathbf{Q}\mathbf{R}_i\big)^2\Big] \\
&= \mathbf{a}_i E\Big[\mathbf{Q}\mathbf{a}_i^H\mathbf{a}_i\mathbf{Q}\Big]\mathbf{a}_i^H + 2b_i^2\mathbf{a}_i E\Big[\mathbf{Q}\mathbf{a}_i^H\ \mathrm{tr}(\mathbf{Q}\mathbf{R}_i)\Big] \\
&\quad + 2b_i^2\mathbf{a}_i E\Big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\Big]\mathbf{a}_i^H + b_i^4\Big\{\mathrm{tr}\Big(E\big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\big]\mathbf{R}_i\Big) + E\Big[\big(\mathrm{tr}(\mathbf{Q}\mathbf{R}_i)\big)^2\Big]\Big\}.
\end{aligned} \tag{C.12}$$

In (C.12) several terms are of the form $E[\mathbf{Q}\mathbf{B}\mathbf{Q}]$. Hence using $\mathbf{Q} = \mathbf{g}_j^H\mathbf{g}_j$, we write

$$\begin{aligned}
E[\mathbf{Q}\mathbf{B}\mathbf{Q}] &= \mathbf{g}_j^H\mathbf{g}_j\mathbf{B}\mathbf{g}_j^H\mathbf{g}_j \\
&= \Big(\mathbf{a}_j^H + b_j\mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H\Big)\Big(\mathbf{a}_j + b_j\ \hat{\mathbf{h}}_j\mathbf{R}_j^{1/2}\Big)\mathbf{B}\Big(\mathbf{a}_j^H + b_j\mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H\Big)\Big(\mathbf{a}_j + b_j\ \hat{\mathbf{h}}_j\mathbf{R}_j^{1/2}\Big) \\
&= \Big(\mathbf{a}_j^H\mathbf{a}_j + \mathbf{a}_j^H b_j\ \hat{\mathbf{h}}_j\mathbf{R}_j^{1/2} + b_j\mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H\mathbf{a}_j + b_j\mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H b_j\ \hat{\mathbf{h}}_j\mathbf{R}_j^{1/2}\Big)\mathbf{B} \\
&\quad \Big(\mathbf{a}_j^H\mathbf{a}_j + \mathbf{a}_j^H b_j\ \hat{\mathbf{h}}_j\mathbf{R}_j^{1/2} + b_j\mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H\mathbf{a}_j + b_j\mathbf{R}_j^{1/2}\ \hat{\mathbf{h}}_j^H b_j\ \hat{\mathbf{h}}_j\mathbf{R}_j^{1/2}\Big).
\end{aligned} \tag{C.13}$$

After some manipulations we get

$$\begin{aligned}
E[\mathbf{Q}\mathbf{B}\mathbf{Q}] &= \mathbf{a}_j^H\mathbf{a}_j\mathbf{B}\mathbf{a}_j^H\mathbf{a}_j + b_j^2\mathbf{a}_j^H\mathbf{a}_j\mathbf{B}\mathbf{R}_j + b_j^2\mathrm{tr}(\mathbf{B}\mathbf{R}_j)\mathbf{a}_j^H\mathbf{a}_j \\
&\quad + b_j^2\mathbf{a}_j\mathbf{B}\mathbf{a}_j^H\mathbf{R}_j + b_j^2\mathbf{R}_j\mathbf{B}\mathbf{a}_j^H\mathbf{a}_j + b_j^4\mathbf{R}_j\mathbf{B}\mathbf{R}_j + b_j^4\mathrm{tr}(\mathbf{B}\mathbf{R}_j)\mathbf{R}_j.
\end{aligned} \tag{C.14}$$

Now rewrite the terms in (C.12) so that they are in [**QBQ**] form. For the first term $\mathbf{B} = \mathbf{a}_i^H \mathbf{a}_i$. The Second term of (C.12) can be written as

$$\begin{aligned} E\Big[\mathbf{a}_i \mathbf{Q} \mathbf{a}_i^H \ \mathrm{tr}(\mathbf{Q}\mathbf{R}_i)\Big] &= E\Big[\mathbf{a}_i \mathbf{g}_j^H \mathbf{g}_j \mathbf{a}_i^H \mathbf{g}_j \mathbf{R}_i \mathbf{g}_j^H\Big] \\ &= E\Big[\mathbf{a}_i \mathbf{g}_j^H \mathbf{g}_j \mathbf{R}_i \mathbf{g}_j^H \mathbf{g}_j \mathbf{a}_i^H\Big] \\ &= \mathbf{a}_i E\Big[\mathbf{g}_j^H \mathbf{g}_j \mathbf{R}_i \mathbf{g}_j^H \mathbf{g}_j\Big] \mathbf{a}_i^H, \end{aligned} \tag{C.15}$$

where $\mathbf{B} = \mathbf{R}_i$. Similarly, the third and fourth terms of (C.12) can be transformed as $\mathbf{B} = \mathbf{R}_i$. Finally, the 5th term can be transformed as

$$\begin{aligned} E\Big[\big(\mathrm{tr}(\mathbf{Q}\mathbf{R}_i)\big)^2\Big] &= E\Big[\mathrm{tr}(\mathbf{g}_j^H \mathbf{g}_j \mathbf{R}_i)^2\Big] \\ &= E\Big[(\mathbf{g}_j \mathbf{R}_i \mathbf{g}_j^H)^2\Big] \\ &= E\Big[(\mathbf{g}_j \mathbf{R}_i \mathbf{g}_j^H)(\mathbf{g}_j \mathbf{R}_i \mathbf{g}_j^H)\Big] \\ &= \mathrm{tr}\Big(\mathbf{R}_i \big\{E[\mathbf{g}_j^H \mathbf{g}_j \mathbf{R}_i \mathbf{g}_j^H \mathbf{g}_j]\big\}\Big), \end{aligned} \tag{C.16}$$

which is also in [**QBQ**] form, where $\mathbf{B} = \mathbf{R}_i$. Thus, (C.12) can be rewritten as

$$\begin{aligned} E\big[(\mathbf{g}_i \mathbf{Q} \mathbf{g}_i^H)^2\big] &= \mathbf{a}_i E\Big[\mathbf{Q}\mathbf{a}_i^H \mathbf{a}_i \mathbf{Q}\Big]\mathbf{a}_i^H + 2b_i^2 \mathbf{a}_i E\Big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\Big]\mathbf{a}_i^H + 2b_i^2 \mathbf{a}_i E\Big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\Big]\mathbf{a}_i^H \\ &\quad + b_i^4 \mathrm{tr}\Big(E\big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\big]\mathbf{R}_i\Big) + b_i^4 \mathrm{tr}\Big(E\big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\big]\mathbf{R}_i\Big) \\ &= \mathbf{a}_i E\Big[\mathbf{Q}\mathbf{a}_i^H \mathbf{a}_i \mathbf{Q}\Big]\mathbf{a}_i^H + 4b_i^2 \mathbf{a}_i E\Big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\Big]\mathbf{a}_i^H + 2b_i^4 \mathrm{tr}\Big(E\big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\big]\mathbf{R}_i\Big). \end{aligned} \tag{C.17}$$

Now using $\mathbf{Q} = \mathbf{g}_j^H \mathbf{g}_j$ and (C.14) in every term of (C.17), where for the first term

$\mathbf{B} = \mathbf{a}_i^H \mathbf{a}_i$ and for all other 4 terms $\mathbf{B} = \mathbf{R}_i$, gives

$$
\begin{aligned}
E\big[(\mathbf{g}_i \mathbf{Q} \mathbf{g}_i^H)^2\big] &= \mathbf{a}_i E\Big[\mathbf{Q}\mathbf{a}_i^H \mathbf{a}_i \mathbf{Q}\Big]\mathbf{a}_i^H + 4b_i^2 \mathbf{a}_i E\Big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\Big]\mathbf{a}_i^H + 2b_i^4 \mathrm{tr}\Big(E\big[\mathbf{Q}\mathbf{R}_i\mathbf{Q}\big]\mathbf{R}_i\Big) \\
&= \mathbf{a}_i \Big\{ \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j + b_j^2 \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H \mathbf{a}_i \mathbf{R}_j + b_j^2 \mathrm{tr}(\mathbf{a}_i^H \mathbf{a}_i \mathbf{R}_j) \mathbf{a}_j^H \mathbf{a}_j \\
&+ b_j^2 \mathbf{a}_j \mathbf{a}_i^H \mathbf{a}_i \mathbf{a}_j^H \mathbf{R}_j + b_j^2 \mathbf{R}_j \mathbf{a}_i^H \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j + b_j^4 \mathbf{R}_j \mathbf{a}_i^H \mathbf{a}_i \mathbf{R}_j + b_j^4 \mathrm{tr}(\mathbf{a}_i^H \mathbf{a}_i \mathbf{R}_j) \mathbf{R}_j \Big\} \mathbf{a}_i^H \\
&+ 4b_i^2 \mathbf{a}_i \Big\{ \mathbf{a}_j^H \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H \mathbf{a}_j + b_j^2 \mathbf{a}_j^H \mathbf{a}_j \mathbf{R}_i \mathbf{R}_j + b_j^2 \mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) \mathbf{a}_j^H \mathbf{a}_j \\
&+ b_j^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H \mathbf{R}_j + b_j^2 \mathbf{R}_j \mathbf{R}_i \mathbf{a}_j^H \mathbf{a}_j + b_j^4 \mathbf{R}_j \mathbf{R}_i \mathbf{R}_j + b_j^4 \mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) \mathbf{R}_j \Big\} \mathbf{a}_i^H \\
&+ 2b_i^4 \mathrm{tr}\Big\{ \mathbf{a}_j^H \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{R}_i + b_j^2 \mathbf{a}_j^H \mathbf{a}_j \mathbf{R}_i \mathbf{R}_j \mathbf{R}_i + b_j^2 \mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) \mathbf{a}_j^H \mathbf{a}_j \mathbf{R}_i \\
&+ b_j^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H \mathbf{R}_j \mathbf{R}_i + b_j^2 \mathbf{R}_j \mathbf{R}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{R}_i + b_j^4 \mathbf{R}_j \mathbf{R}_i \mathbf{R}_j \mathbf{R}_i + b_j^4 \mathbf{R}_j^{1/2} \mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) \mathbf{R}_j^{1/2} \mathbf{R}_i \Big\}.
\end{aligned}
\tag{C.18}
$$

Finally, we need $E\big[(\mathbf{g}_i \mathbf{Q} \mathbf{g}_i^H)\big]^2$ to compute the variance. We know that

$$
E\big[(\mathbf{g}_i \mathbf{Q} \mathbf{g}_i^H)\big] = \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H + b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H + b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H + b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j), \tag{C.19}
$$

and can compute

$$
\begin{aligned}
E\big[(\mathbf{g}_i \mathbf{Q} \mathbf{g}_i^H)\big]^2 &= \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H + \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H \\
&+ \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H + \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) \\
&+ b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H + b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H \\
&+ b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H + b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) \\
&+ b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H + b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H \\
&+ b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H + b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) \\
&+ b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) \mathbf{a}_i \mathbf{a}_j^H \mathbf{a}_j \mathbf{a}_i^H + b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) b_j^2 \mathbf{a}_i \mathbf{R}_j \mathbf{a}_i^H \\
&+ b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) b_i^2 \mathbf{a}_j \mathbf{R}_i \mathbf{a}_j^H + b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j) b_i^2 b_j^2 \,\mathrm{tr}(\mathbf{R}_i \mathbf{R}_j).
\end{aligned}
\tag{C.20}
$$

Now, we can compute $\mathrm{Var}[P_{ij}]$ from (C.18) and (C.20) as

$$\begin{aligned}
\mathrm{Var}[P_{ij}] &= E\big[(\mathbf{g}_i\mathbf{Q}\mathbf{g}_i^H)^2\big] - E\big[(\mathbf{g}_i\mathbf{Q}\mathbf{g}_i^H)\big]^2 \\
&= \mathbf{a}_i\Big\{b_j^2\mathbf{a}_j\mathbf{a}_i^H\mathbf{a}_i\mathbf{a}_j^H\mathbf{R}_j + b_j^2\mathbf{R}_j\mathbf{a}_i^H\mathbf{a}_i\mathbf{a}_j^H\mathbf{a}_j + b_j^4\mathbf{R}_j\mathbf{a}_i^H\mathbf{a}_i\mathbf{R}_j\Big\}\mathbf{a}_i^H \\
&+ b_i^2\mathbf{a}_i\Big\{2\mathbf{a}_j^H\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H\mathbf{a}_j + 4b_j^2\mathbf{a}_j^H\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j + 2b_j^2\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j)\mathbf{a}_j^H\mathbf{a}_j \\
&+ 2b_j^2\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H\mathbf{R}_j + 4b_j^2\mathbf{R}_j\mathbf{R}_i\mathbf{a}_j^H\mathbf{a}_j + 4b_j^4\mathbf{R}_j\mathbf{R}_i\mathbf{R}_j + 2b_j^4\mathbf{R}_j^{1/2}\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j)\mathbf{R}_j^{1/2}\Big\}\mathbf{a}_i^H \\
&+ b_i^4\mathrm{tr}\Big\{\mathbf{a}_j^H\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H\mathbf{a}_j\mathbf{R}_i + 2b_j^2\mathbf{a}_j^H\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j\mathbf{R}_i + 2b_j^2\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H\mathbf{R}_j\mathbf{R}_i \\
&+ 2b_j^2\mathbf{R}_j\mathbf{R}_i\mathbf{a}_j^H\mathbf{a}_j\mathbf{R}_i + 2b_j^4\mathbf{R}_j\mathbf{R}_i\mathbf{R}_j\mathbf{R}_i + b_j^4\mathbf{R}_j^{1/2}\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j)\mathbf{R}_j^{1/2}\mathbf{R}_i\Big\}.
\end{aligned} \tag{C.21}$$

Simplifying (C.21), we get

$$\begin{aligned}
\mathrm{Var}[P_{ij}] &= b_j^2\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H|\mathbf{a}_i\mathbf{a}_j^H|^2 + b_j^2\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H|\mathbf{a}_i\mathbf{a}_j^H|^2 + b_j^4(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H)^2 \\
&+ 2b_i^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)|\mathbf{a}_i\mathbf{a}_j^H|^2 + 4b_i^2b_j^2(\mathbf{a}_i\mathbf{a}_j^H)(\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j\mathbf{a}_i^H) + 2b_i^2b_j^2|\mathbf{a}_i\mathbf{a}_j^H|^2\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) \\
&+ 2b_i^2b_j^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H) + 2b_i^2b_j^2(\mathbf{a}_j\mathbf{a}_i^H)(\mathbf{a}_i\mathbf{R}_j\mathbf{R}_i\mathbf{a}_j^H) + 4b_i^2b_j^4\mathbf{a}_i(\mathbf{R}_j\mathbf{R}_i\mathbf{R}_j)\mathbf{a}_j^H \\
&+ 2b_i^2b_j^4(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H)\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) + b_i^4(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)^2 + 2b_i^4b_j^2\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j\mathbf{R}_i\mathbf{a}_j^H \\
&+ 2b_i^4b_j^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) + 2b_i^4b_j^2\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j\mathbf{R}_i\mathbf{a}_j^H + 2b_i^4b_j^4\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j\mathbf{R}_i\mathbf{R}_j) \\
&+ b_i^4b_j^4\mathrm{tr}[(\mathbf{R}_i\mathbf{R}_j)]^2.
\end{aligned} \tag{C.22}$$

Hence, the variance can be written as

$$\begin{aligned}
\mathrm{Var}[P_{ij}] &= 2b_j^2|\mathbf{a}_i\mathbf{a}_j^H|^2\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H + b_i^4(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)^2 \\
&+ b_j^4(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H)^2 + 2b_i^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)|\mathbf{a}_i\mathbf{a}_j^H|^2 \\
&+ 8b_i^2b_j^2Re\Big\{(\mathbf{a}_i\mathbf{a}_j^H)(\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j\mathbf{a}_i^H)\Big\} \\
&+ 2b_i^2b_j^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H) + 4b_i^2b_j^4\mathbf{a}_i\mathbf{R}_j\mathbf{R}_i\mathbf{R}_j\mathbf{a}_i^H \\
&+ 4b_i^4b_j^2\mathbf{a}_j\mathbf{R}_i\mathbf{R}_j\mathbf{R}_i\mathbf{a}_j^H + 2b_i^2b_j|\mathbf{a}_i\mathbf{a}_j^H|^2\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) \\
&+ 2b_i^2b_j^4(\mathbf{a}_i\mathbf{R}_j\mathbf{a}_i^H)\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) + 2b_i^4b_j^2(\mathbf{a}_j\mathbf{R}_i\mathbf{a}_j^H)\mathrm{tr}(\mathbf{R}_i\mathbf{R}_j) \\
&+ 2b_i^4b_j^4\mathrm{tr}[(\mathbf{R}_i\mathbf{R}_j)^2] + b_i^4b_j^4\mathrm{tr}[(\mathbf{R}_i\mathbf{R}_j)]^2.
\end{aligned} \tag{C.23}$$